



# UNIVERSIDAD DE LA RIOJA

## TRABAJO FIN DE ESTUDIOS

Título

ETL sobre base de datos educativa

Autor/es

GERMÁN CÁMARA JIMÉNEZ

Director/es

CÉSAR DOMÍNGUEZ PÉREZ y EMILIO RODRÍGUEZ PRIEGO ,

Facultad

Facultad de Ciencia y Tecnología

Titulación

Grado en Ingeniería Informática

Departamento

MATEMÁTICAS Y COMPUTACIÓN

Curso académico

2018-19



***ETL sobre base de datos educativa***, de GERMÁN CÁMARA JIMÉNEZ  
(publicada por la Universidad de La Rioja) se difunde bajo una Licencia Creative  
Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 Unported.  
Permisos que vayan más allá de lo cubierto por esta licencia pueden solicitarse a los  
titulares del copyright.



# **UNIVERSIDAD DE LA RIOJA**

**Facultad de Ciencia y Tecnología**

## **TRABAJO FIN DE GRADO**

**Grado en Ingeniería Informática**

### **ETL SOBRE BASE DE DATOS EDUCATIVA**

Realizado por:

Germán Cámara Jiménez

Tutelado por:

César Domínguez Pérez

Emilio Rodríguez Priego

**Logroño, Julio, 2019**

## Índice

<b>Resumen:</b>	<b>3</b>
<b>Abstract:</b>	<b>3</b>
<b>1. Introducción</b>	<b>4</b>
1.1. Contexto	4
1.2. Tecnologías	5
1.3. Alcance	6
1.3.1. Requisitos	6
1.3.2. Exclusiones	6
1.3.3. Entorno	6
1.4. Gestión del proyecto	7
1.4.1. Metodología	7
1.4.2. Estructura de descomposición de tareas (EDT)	8
1.4.3. Descripción de tareas	9
1.4.4. Plan de comunicación	10
1.4.5. Plan de contingencia	10
1.4.6. Calidad	11
<b>2. Capítulo 1: estudio de la base de datos y tecnologías</b>	<b>12</b>
2.1. Análisis de tecnologías	12
2.2. Base de datos RACIMA	13
2.2.1. Estudio previo	14
2.2.2. Primer estudio de la base de datos de RACIMA	15
<b>3. Capítulo 2: diseño de hechos y dimensiones</b>	<b>30</b>
<b>4. Capítulo 3: comparaciones entre alumnos con fracaso y sin fracaso</b>	<b>35</b>
<b>5. Capítulo 4: implementación de los hechos y dimensiones</b>	<b>43</b>
5.1. Tablas de dimensiones	44
5.2. Tablas de hechos	46
<b>6. Creación de un cubo</b>	<b>49</b>
<b>7. Comparativa entre el tiempo estimado y el real</b>	<b>52</b>
<b>8. Conclusiones</b>	<b>54</b>
<b>9. Bibliografía</b>	<b>55</b>

## Resumen

Este Trabajo de Fin de Grado (TFG) ha sido propuesto por Francisco Martínez Martín, responsable del proyecto del sistema educativo de RACIMA, y Óscar Alonso Hernández, jefe del Área de Infraestructuras Territoriales que incluye todos los sistemas de información relacionados con el sistema educativo a nivel preuniversitario, con el objetivo de poder analizar las causas del fracaso escolar y poder detectar posibles casos a tiempo y así prevenirlo.

Para llevar a cabo este objetivo, se va a emplear la base de datos del sistema de educación de La Rioja llamada RACIMA.

El punto principal consistirá en analizar dicha base de datos, y así, por un lado, poder entender correctamente cómo se estructura y cómo se encuentra organizada, y por el otro, ver los datos disponibles en ella. Posteriormente, tras realizar un análisis exhaustivo de las posibles causas del fracaso escolar, se llevará a cabo sobre la base de datos un proceso de ETL (Extraction, Transformation, Load) para facilitar el estudio de dichas causas.

## Abstract

This Final Degree Project has been proposed by Francisco Martínez Martín, head of the RACIMA education system project, and Óscar Alonso Hernández, head of the Territorial Infrastructure Area, which includes all the information systems related to the education system at the pre-university level, with the aim of being able to analyze the causes of school failure and to detect possible cases in time and prevent it.

To carry out this objective, the database of the education system of La Rioja called RACIMA will be used.

The main point will be to analyze this database, and thus, to understand correctly how it is structured and how it is organized, and to see the data available in it. Subsequently, after an exhaustive analysis of the possible causes of school failure, an ETL process will be carried out on the database to facilitate the study of these causes.

## 1. Introducción

En este documento se recogerán los puntos más importantes del estudio de la base de datos de la plataforma educativa RACIMA y de su posterior ejercicio de ETL. El objetivo final es analizar la base de datos de la plataforma.

### 1.1. Contexto

En el Gobierno de La Rioja se dispone de una base de datos para almacenar toda la información referente a la educación, y se trata de la plataforma educativa llamada RACIMA. En ella se contienen una gran cantidad de datos sobre los diferentes centros y alumnos de toda La Rioja. Esta plataforma permite realizar todo tipo de tareas y actividades que tengan relación con la gestión académica de los centros, así como realizar un seguimiento educativo de cada alumno. Asimismo, puede ser utilizada por un gran número de personas, que engloban desde personas con cargos directivos o profesores, hasta padres y alumnos.

RACIMA surgió a partir de la plataforma empleada en Andalucía llamada SENECA, con el objetivo de implantar en La Rioja un sistema de educación inspirada en la andaluza.

Esta plataforma posee 3 módulos: gestión de centros, seguimiento educativo y secretaría virtual, que se detallan a continuación:

**Gestión de centros:** RACIMA posee herramientas para facilitar la gestión de los centros y para poder centralizar su información. Asimismo contiene aplicaciones para poder gestionar cualquier actividad diaria, como la jornada escolar, el calendario escolar, admisiones, matrículas, licencias y profesor entre otras.

**Seguimiento educativo:** Gracias a este módulo, los padres pueden hacer un seguimiento de la vida escolar de sus hijos, permitiéndoles tener acceso a sus horarios, las faltas a clase, trabajos y tareas. Por otro lado, a los alumnos se les permite consultar sus calificaciones y las fechas de los controles. Por último, este módulo tiene un espacio para las AMPAs, que permite potenciar su participación en los centros.

**Secretaría virtual:** Este módulo permite a los padres efectuar trámites administrativos desde casa, sin necesidad de desplazarse al centro. Los equipos directivos pueden realizar trámites internos con la Consejería de Educación.

La base de datos de RACIMA consta de más de un millar de tablas, de las cuales se han empleado en este trabajo 1170 y están desplegadas en Oracle 11g. Algunas de estas tablas no contienen datos, debido al proceso de herencia a partir de SENECA, donde se incorporaron todas las existentes, quedando algunas tablas sin utilidad, por lo que se decidió no usarlas en RACIMA, dejándolas en blanco.

El propósito de este TFG y del cliente para el cual se realiza este trabajo es analizar la base de datos, para poder ver las posibles causas del fracaso escolar. Para ello, se analiza a los alumnos con fracaso, buscando las características que tienen en común, y se comparan con los que no tienen fracaso, para poder tomar medidas y así reducir dicho fracaso. Este estudio está basado en la percepción que tenemos del fracaso escolar y las causas que lo provocan, poniendo en común con el cliente las posibles fuentes de fracaso escolar para posteriormente estudiarlas.

Particularmente, el fracaso escolar tiene muchas acepciones, y en comunicación con el cliente, se ha decidido para este proyecto que el fracaso escolar se produce cuando un alumno termina 4º de la E.S.O. sin obtener el título acreditativo del mismo.

## 1.2. Tecnologías

En este TFG no se ha realizado un estudio previo de las tecnologías dado que su uso ya ha venido impuesto de antemano. Se han empleado las tecnologías Oracle 11g, SQL Developer y Pentaho, que se detallan a continuación:

- **Oracle 11g:** La base de datos está guardada en Oracle, por tanto, se utilizará Oracle 11g o una versión posterior como sistema gestor de base de datos.
- **SQL Developer:** Como cliente de Oracle, se utilizará la herramienta SQL Developer para la visualización y realización de operaciones sobre la base de datos dado que es el cliente por excelencia de Oracle.

- **Pentaho:** Para la parte de la ETL (Extraction, Transformation, Load), se utilizará Pentaho. En concreto, se emplea Spoon, que es una herramienta open source y que da mayor funcionalidad de manera intuitiva. Destacar que se ha elegido esta herramienta porque es la que utiliza el Gobierno de La Rioja para trabajar sobre RACIMA.

### 1.3. Alcance

Como se ha comentado anteriormente, el objetivo del proyecto es realizar un ejercicio de ETL sobre la base de datos RACIMA para hacer un estudio de las causas del fracaso escolar.

#### 1.3.1. Requisitos

- Se estudiará la base de datos RACIMA, para poder entenderla claramente.
- Se diseñará un Data Warehouse con las dimensiones detectadas durante la fase de estudio, de las cuáles se sacarán posibles hechos.
- Se diseñará otra distribución de las tablas de la base de datos para disponerlas en forma de estrella (tablas de hechos y dimensiones).
- Se procederá a realizar un ejercicio de ETL sobre la base de datos RACIMA para pasar los datos de la base de datos original al nuevo diseño.

#### 1.3.2. Exclusiones

- No se realizará *business intelligence* sobre el resultado final.

#### 1.3.3. Entorno

- Como se ha indicado, se utilizará SQL Developer como cliente de Oracle para realizar el trabajo.
- Se empleará la herramienta Pentaho para la realización del proceso de ETL sobre la base de datos. Para ello, se necesitará un estudio previo de esta herramienta.



## 1.4. Gestión del proyecto

La fecha de finalización del proyecto se prevé sobre finales del mes de junio, en el primer depósito de TFGs. Para ello emplearé 5 horas diarias en el trabajo. El trabajo comenzará el lunes 4 de febrero y se prevé acabarlo sobre mediados de junio, dando margen por si hubiese algún contratiempo y dejando tiempo para los exámenes.

Las primeras semanas se emplearán en realizar reuniones con el cliente para establecer el objetivo final del trabajo. Tras esto, la siguiente labor será realizar el estudio de la base de datos mediante la herramienta SQL Developer.

Tras la finalización del análisis de la base de datos, lo siguiente será estudiar a los alumnos que hayan sufrido fracaso escolar, para analizar las posibles causas, que luego se traducirían en un conjunto de tablas en forma de estrella.

En el momento en que ya se haya determinado qué hechos son importantes, se realizará el proceso de ETL construyendo las tablas de hechos y dimensiones de los hechos que se consideren importantes utilizando la herramienta Pentaho.

### 1.4.1. Metodología

Dado que el trabajo se va a realizar por etapas, la metodología elegida es la de cascada o waterfall. Las etapas en las que se ha dividido el proyecto son las siguientes:

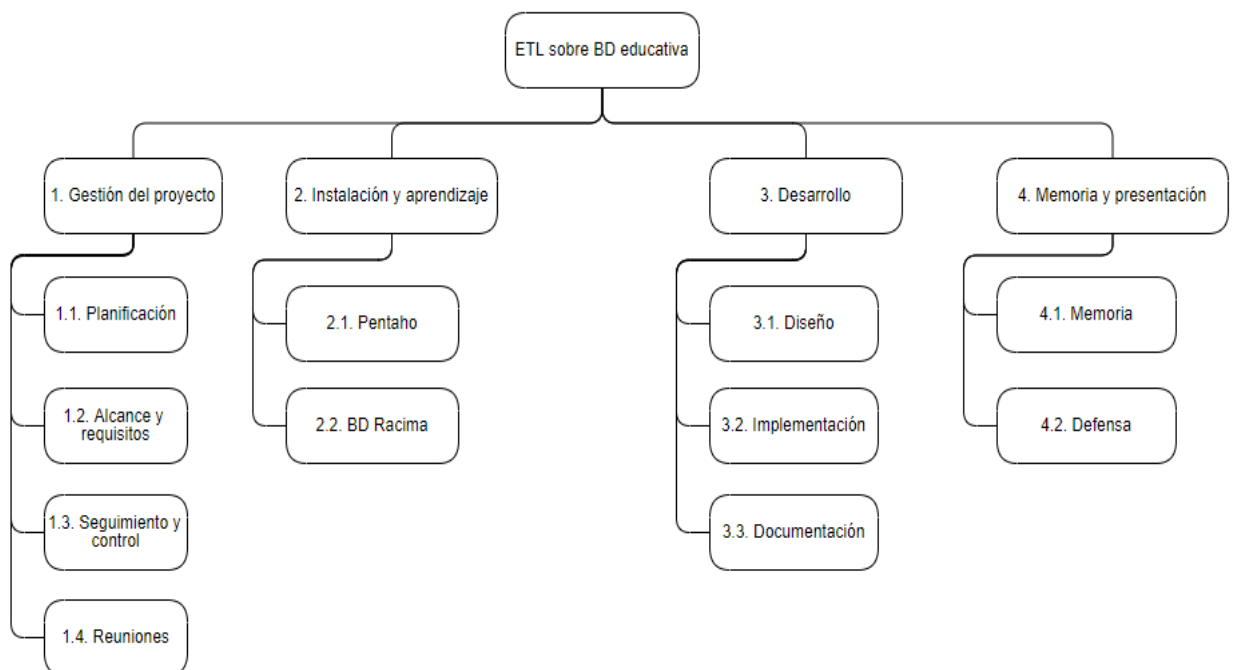
1. **Análisis de la base de datos y tecnologías:** Consiste en realizar el estudio de la base de datos. Para ello, se estudiará cada una de las tablas que la componen y su estructura interna. Al emplear tecnologías nuevas, habrá que estudiarlas previamente, para poder hacer un uso correcto de ellas.
2. **Diseño de hechos y dimensiones:** A partir del estudio de la fase anterior, se procederá a diseñar posibles hechos en forma de estrella (tabla de hechos y tablas de dimensiones).
3. **Comparaciones entre alumnos con fracaso escolar y sin fracaso:** En esta etapa se separarán los alumnos en dos grupos. Por un lado, los que tienen fracaso escolar (no obtienen título) y por el otro, los que no han

fracasado (obtienen título). La idea de esta etapa es sacar a la luz posibles patrones que indiquen causas de fracaso escolar.

4. **Implementación de los hechos y dimensiones:** Como paso final, se implementarán los hechos y dimensiones utilizando Pentaho.

#### 1.4.2. Estructura de descomposición de tareas (EDT)

A continuación, en la Figura 1.1, se puede observar la estructura de descomposición de tareas.



**Figura 1.1.** Estructura de descomposición de tareas

### 1.4.3. Descripción de tareas

En la tabla siguiente se describen detalladamente cada una de las tareas reflejadas en el EDT anterior (Figura 1.1) junto con la estimación de tiempo.

Descripción tareas	
Nombre paquete	Descripción
1.1. Planificación	Definición de las tareas a realizar, la forma de proceder y las horas estimadas.
1.2. Alcance y requisitos	Especificación de cuál es el objetivo y lo que se pretende conseguir y lo que no.
1.3. Seguimiento y control	Recoge los cambios que ha ido sufriendo el proyecto a lo largo de su realización e incluye comparaciones con lo planificado.
1.4. Reuniones	Información sobre las reuniones mantenidas tanto con el cliente como con los tutores.
<b>Tiempo estimado</b>	<b>70 horas</b>
2.1. Pentaho	Lectura de manuales sobre Pentaho además de su instalación y la realización de pruebas siguiendo tutoriales.
2.2. BD Racima	Instalación de la base de datos y estudio de las tablas que la componen y de su estructura, tomando anotaciones para aligerar futuras consultas sobre ella.
<b>Tiempo estimado</b>	<b>90 horas</b>
3.1. Diseño	Diseños realizados después del estudio de la base de datos para encontrar posibles hechos de interés para el proyecto.
3.2. Implementación	Construcción de los hechos en formato de estrella utilizando la herramienta Pentaho siguiendo los diseños que finalmente se hayan seleccionado.
3.3. Documentación	Información sobre las tablas de la base de datos Racima. Incluye el

	nombre de la tabla, lo que representa y los campos considerados más importantes.
<b>Tiempo estimado</b>	<b>90 horas</b>
4.1. Memoria	Redacción de la memoria del proyecto.
4.2. Defensa	Presentación del proyecto acompañado de un powerpoint.
<b>Tiempo estimado</b>	<b>50 horas</b>
<b>Tiempo total</b>	<b>300 horas</b>

#### 1.4.4. Plan de comunicación

A lo largo del desarrollo del proyecto, se realizarán numerosas reuniones tanto con el cliente como con los tutores del TFG. La idea principal es que ambas partes estén al corriente de los avances en el proyecto. Para ello, habrá dos tipos de reuniones:

- 1) **Comunicación con el cliente:** La comunicación con el cliente será vía correo electrónico, y vía oral cuando se organicen reuniones. Si hubiese algún cambio de última hora en las reuniones, la comunicación sería por teléfono móvil.
- 2) **Comunicación con el tutor:** La comunicación con el tutor de la universidad será mediante correo electrónico, utilizando para ello el correo electrónico personal, y vía oral cuando se organicen reuniones.

#### 1.4.5. Plan de contingencia

Para evitar la pérdida de información, se realizará una copia en local y otra en One Drive de los archivos del cliente. De igual manera, la documentación del trabajo se alojará en Drive para tenerlo todo en la nube.

En caso de que el cliente no pueda organizar una reunión, debido a que se puede avanzar de fase hasta finalizar la anterior, se dedicará el tiempo a realizar la memoria del proyecto. Si el cliente pide realizar cambios, se empleará el tiempo estimado para las fases finales del proyecto.

#### 1.4.6. Calidad

Con el objetivo de conseguir el producto que quiere el cliente y evitar posibles cambios a mitad de proyecto, se entregará al cliente los diseños realizados así como futuras ideas para su valoración y así pueda expresar si está de acuerdo o no, con lo mostrado.

## 2. Capítulo 1: estudio de la base de datos y tecnologías

En este capítulo se estudiarán las tablas que componen la base de datos, con el objetivo de ver qué información almacenan y si será útil en un futuro. Este estudio sirve también para familiarizarme con la terminología usada para dar nombre a las tablas y a los campos de las mismas. A su vez será necesario familiarizarse con las tecnologías que utilizaré para realizar el trabajo, como son SQL Developer y Pentaho, siendo esta última una total novedad para mi experiencia tecnológica.

### 2.1. Análisis de tecnologías

Pentaho es una herramienta de Business Intelligence de código abierto (open source) que ofrece soluciones para la gestión y análisis de la información.

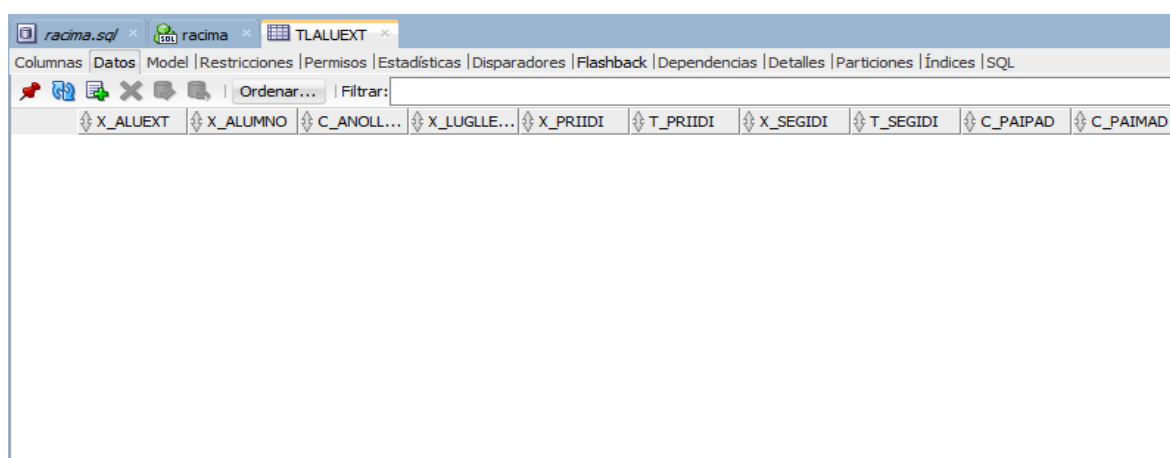
Posee varios módulos:

- **Pentaho reporting:** Recoge y muestra informes de la empresa pudiendo ser visualizados en PDFs, HTML, Excel...
- **Pentaho Data Integration:** A esta funcionalidad se le llama también Kettle o Spoon, y es la herramienta que se utilizará en este TFG. Transforma e integra los datos entre sistemas de información y los datamarts que componen el sistema BI.
- **Pentaho analysis:** Este módulo permite consultar y analizar la información del negocio de forma interactiva.
- **Pentaho Data Mining:** Permite descubrir patrones de comportamiento de la información.
- **Sparkl:** Permite crear plugins personalizados orientados a determinadas funcionalidades y objetivos.
- **Pentaho Big Data:** Ofrece las herramientas necesarias para poder analizar grandes volúmenes de datos.

## 2.2. Base de datos RACIMA

La base de datos de RACIMA es muy amplia. Tal y como se ha comentado en el capítulo anterior, han sido facilitadas sólo 1170 tablas del total. Todas las tablas tienen nombres especiales según lo que contienen. Empiezan por **TL**, que significa **tabla**, y continúan con sus nombres completos en caso de ser una única palabra, por ejemplo, la tabla de alumnos sería TLALUMNOS, o con abreviaciones utilizando las tres primeras letras de cada palabra en caso de estar compuestas por más de una palabra, por ejemplo, la tabla matrícula de alumno sería TLMATALU. Los campos que forman las tablas utilizan un sistema de nombres, que consiste en una letra que indica el tipo de contenido, guion bajo y el nombre del campo. Los que empiezan por **X\_** significa que esa columna es el código interno numérico, los que empiezan por **F\_** son fechas (dates), los que empiezan por **D\_** son descripciones largas y los que empiezan por **S\_** son las descripciones cortas, los que empiezan por **T\_** son campos de texto y, finalmente, los que empiezan por **C\_** son campos con códigos visibles alfabéticos. Hay tablas que empiezan por **ADM\_**, por **ACTA\_** y por **V\_**, pero no son relevantes en el estudio de la base de datos.

Tal y como se ha comentado anteriormente, RACIMA tuvo su origen en la aplicación SENECA de Andalucía. Debido a que el proceso de herencia incorporó todas las tablas, existen tablas en la base de datos de RACIMA que no contienen dato alguno, debido a que esa misma tabla contenía datos en la base de datos de SENECA, pero se decidió que no era útil en RACIMA y se dejó vacía como se muestra en la Figura 2.1.



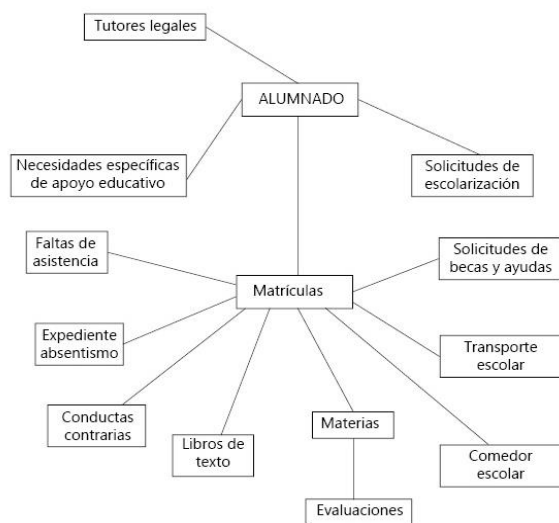
**Figura 2.1.** Tabla vacía de la base de datos de RACIMA

Asimismo, existen otras tablas vacías, pero el hecho de que estén así es debido a que se ha proporcionado la base de datos anonimizada y los datos psicológicos y clínicos de los alumnos y profesores no se han incorporado. Por el mismo motivo, los nombres y apellidos de las personas han sido modificados aleatoriamente.

### 2.2.1. Estudio previo

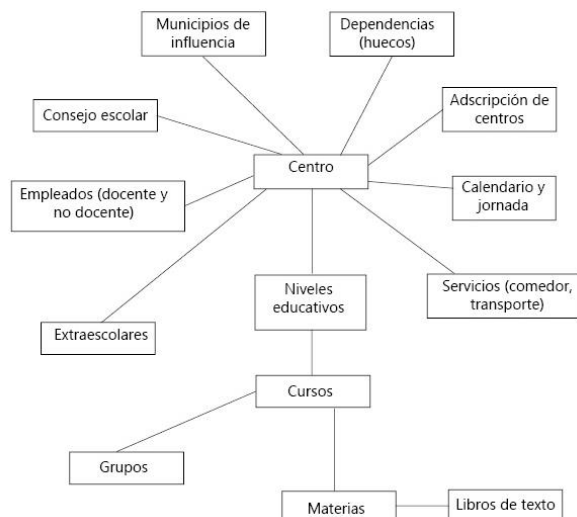
En mi primera reunión con el cliente, me estuvo preguntando cuál creía yo que era el propósito del proyecto y cuestiones sobre el proceso de ETL para comprobar que yo no presentaba falta de conceptos sobre el tema. Más tarde, me dijo que no podían darme acceso directamente a la base de datos de RACIMA. Esto se debía a que tenían que realizar un proceso en el que se iban a seleccionar las tablas que podían darme, debido a que las tablas que contenían información crítica de las personas (detalles médicos, psicológicos...). Además, tenían que anonimizar la base de datos, es decir, cambiar todos los datos que podrían permitir identificar a una persona (Nombre, apellidos, DNI...) y, para ello, tenían que pasar las tablas por un proceso de anonimización en el que de forma aleatoria iban a cambiar dichos datos.

Inicialmente el cliente en unas reuniones previas adjuntó unas vistas (Figura 2.2, Figura 2.3 y Figura 2.4) que resumían la base de datos, con el objetivo de facilitar una idea previa del material con el que se iba a contar, mientras se preparaba y anonimizaba la base de datos, para poder ser importada y trabajar con ella, tal y como se ha comentado en el apartado anterior y así cumplir con la Ley de Protección de Datos.

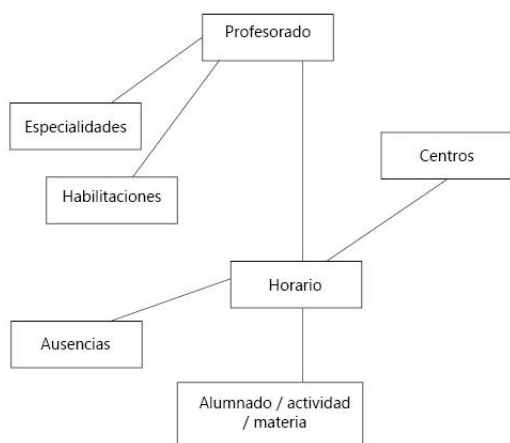


**Figura 2.2** Vistas relacionadas con las matrículas





**Figura 2.3** Vistas relacionadas con los centros



**Figura 2.4** Vistas relacionadas con el horario del profesor

### 2.2.2. Primer estudio de la base de datos de RACIMA

Hay de destacar que al principio hubo una serie de problemas para desplegar la base de datos en el ordenador. Cuando se hizo el proceso de anonimización y de exportación, no sólo se exportaron las tablas, sino todo el esquema y los tablespaces que se usan, de tal manera que al importarla, el script de importación busca un entorno similar en el ordenador (busca los tablespaces que se utilizan en la base de datos de RACIMA original). Al no encontrarlos aparecía un mensaje diciendo que el script de importación estaba dando errores. Al no encontrar la solución se optó por acudir a uno de los responsables de la exportación de la base de datos para que ayudara a corregir los errores y, tras construir en el ordenador un entorno similar al que utiliza el cliente, se pudo importar la base de datos al ordenador.

Cuando se consiguió desplegar la base de datos, se procedió a analizar el contenido de las tablas. En un principio se revisaron las tablas que cuyo nombre empieza por **ACTA\_** y **ADM\_**, pero luego se observó que no tenían utilidad sobre el proyecto dado que las tablas que contenían datos útiles empezaban por **TL\_**, por lo que se excluyeron del estudio.

Por otro lado, se ha averiguado la funcionalidad de las tablas, ya que cada campo iba acompañado de un comentario descriptivo. Al no existir un manual explicativo de la base de datos, se procedió a revisar una por una las 1170 tablas que componían la base de datos para averiguar cuál era su contenido y lo que representaban. A continuación, se nombran y detallan las tablas que se emplearon tras revisar por primera vez la base de datos (no están todos sus campos, sólo los considerados importantes):

**TLACTEXT:** Tabla de actividades extraescolares. Contiene el nombre de la actividad, las fechas de aprobación, la hora, costes, el medio de transporte, satisfacción del alumnado y comentarios sobre la actividad y los alumnos.

**TL ACTIVIDADES:** Tabla de actividades que puede realizar un profesor (recreo, tutoría, impartir clase...). Contiene el ID de actividad, una descripción de la misma, si la actividad está activa o no, si es tutoría y si es docencia directa.

**TLADJCONVSOLBECAYU:** Tabla de adjudicación de convocatorias de solicitudes de becas o ayudas. Incluye datos sobre el estado de la adjudicación, el tipo de beca, la fecha de solicitud, certificados de defunción de familiares, si es familia numerosa, monoparental, estado de los padres (divorciados, casados...), niveles de renta, ID de desestimación (en caso de haberla), si el alumno rechaza la adjudicación, alegaciones y documentación de los miembros de la unidad familiar.

**TLALUABS:** Tabla de alumnos absentistas. Contiene el ID de alumno, su ID de matrícula, fecha de apertura y cierre del expediente de absentismo (si es que tiene), el año académico y curso, ID del centro, observaciones, si el absentismo se subsanó mediante intervención policial y si la dirección general inició el expediente sancionador.

**TLALUCEN:** Tabla que contiene la relación entre alumnos y centros. Contiene el ID de centro, nombre del alumno, apellidos, DNI, domicilio, localidad, teléfono, sexo del alumno, la ruta del autobús que utiliza (si lo hace), si usa el servicio de comedor y algunas observaciones.

**TLALUCONCON:** Tabla de alumnos con conductas contrarias. Contiene el ID de alumno, el ID de su matrícula, la fecha en la que ocurrió la conducta contraria, el estado de la corrección de la conducta, si el tutor está de acuerdo con la corrección, flag que indica si la conducta ha sido notificada a los tutores, si ha sido una conducta colectiva, fechas de apertura y cierre del expediente y el grado de participación del alumno en el caso de que sea una conducta contraria colectiva.

**TLALUEXT:** Tabla que contiene la información de los alumnos extranjeros. Contiene el ID de alumno, el año de llegada a España, ID del primer idioma, país del padre y de la madre, si ha estado escolarizado antes y dónde, el número de años escolarizado y el tiempo (años, meses, semanas, horas) que ha estado asistiendo a ATAL (Aulas Temporales de Adaptación Lingüística) para adaptarse lingüísticamente.

**TLALUMNOS:** Tabla donde se almacenan todos los alumnos. Contiene el ID de alumno, el código del país de nacionalidad, el país de nacimiento, el número de identificación escolar, fecha de entrada al sistema educativo, nombre y apellidos, sexo, localidad, municipio, provincia, fecha nacimiento, domicilio, teléfono y el identificador interno de la familia.

**TLALUPARASGHORCLA:** Tabla de asignación de asignaturas a alumnos. Contiene el ID de la tabla, el ID de la materia de la matrícula del alumno asignado, el ID del horario de clase y el ID de matrícula de alumno.

**TLALUPLAREFCEN:** Tabla de los planes de refuerzo del centro a los alumnos. Contiene el ID de alumno, el ID del plan de refuerzo, el ID de matrícula y el ID del grupo del plan de refuerzo.

**TLALUPRUPSI:** Tabla de alumnos que han realizado alguna prueba psicopedagógica. Contiene el ID de alumno, el ID de la prueba, el ID de la matrícula del alumno y la valoración de la prueba.

**TLANTALUTIPEXP:** Tabla del expediente de un alumno. Contiene el ID del tipo de expediente, el ID de alumno, el año académico, el estado genérico de la matrícula (obtiene título, no obtiene título, promociona...), el ID del centro y descripción de los estudios realizados.

**TLARECON:** Tabla que tiene las áreas de conocimiento. Contiene el ID del área de conocimiento (griego, latín, biología y geología, matemáticas...) y una descripción larga y otra corta del área.

**TLASIGIMPHORCLA:** Tabla de asignación de la impartición de asignaturas en el horario de clase. Contiene el ID de la materia genérica impartida en el horario de la clase y el ID de la configuración horaria del aula y del profesor.

**TLASIPROPAD:** Tabla de asistencia del horario del profesor. Contiene el ID del horario del profesor y la fecha de asistencia del profesor a clase.

**TLASIUSERCOM:** Tabla de asistencia al comedor. Contiene el ID de asistencia, el ID del usuario del servicio y la fecha de asistencia al servicio de comedor.

**TLAULPAD:** Tabla de las aulas. Contiene el ID del aula, el ID del centro al que pertenece, el nombre del aula y su capacidad.

**TLAUSENCIAS:** Tabla de las ausencias del profesorado. Contiene el ID del empleado que se ausenta, el ID del motivo de la ausencia, el ID del centro al que falta, las fechas de inicio y fin de las ausencias, el ID del tipo de justificación, el número de días hábiles que falta, la fecha de la comunicación de la ausencia y si la ausencia queda o no justificada.

Merece la pena destacar que en RACIMA, si un alumno no va a clase no se le llama ausencia sino falta de asistencia.

**TLAUSTRAHOR:** tabla de ausencias en el tramo horario. Contiene el ID del tramo horario de una ausencia y el ID de la ausencia de un empleado.

**TLCALIFICACIONES:** Tabla de la definición de las calificaciones, es decir, en la que se definen las notas (de 0 a 4 es suspenso, un 5 es suficiente, un 6 es un bien...). Contiene el ID de la calificación, su descripción, su abreviatura y si con esa calificación se aprueba.

**TLCALPRUACC:** Tabla de calificaciones de las pruebas de acceso. Contiene el ID del alumno evaluado, el ítem evaluado y el ID de la calificación.

**TLCENADS:** Tabla de centros adscritos. Contiene el ID del centro de referencia y el centro adscrito al centro de referencia.

**TLCENADSIES:** Tabla de centros adscritos a sus IES. Contiene el ID del centro adscrito a su IES, el centro de referencia, el centro adscrito al centro de referencia y el año de inicio y fin de la relación de adscripción.

**TLCENAUTENS:** Tabla de autorización de enseñanza. Contiene el ID del centro, el período de autorización de enseñanza, el número de la localidad (si es un 0, significa que es un centro distribuido en varias localidades) y la descripción de la localidad.

**TLCENTROS:** tabla de centros. Contiene el ID del tipo de centro, la fecha de alta y de baja y un flag que indica si está cesado.

**TLCOMALUCONCON:** tabla de comunicados de alumnos con conductas contrarias. Contiene el ID de matrícula, el ID del empleado que registra la conducta, la fecha del comunicado, una descripción breve y otra detallada y el momento en que se produce la conducta contraria.

**TLCOMESC:** Tabla del comedor escolar. Contiene el ID del comedor, el ID de la modalidad de comedor escolar, el ID del centro al que pertenece, la fecha de inicio y fin del servicio de comedor, las horas de inicio y fin del servicio y el número de plazas disponibles.

**TLCONCONCOL:** Tabla de conductas contrarias colectivas. Contiene su ID, la fecha en la que ocurrió y su descripción.

**TLCORCONNEGALU:** Tabla de correcciones de conductas negativas (o contrarias) de un alumno. Contiene la matrícula del alumno y el tipo de conducta contraria.

**TLCURSOACA:** Tabla de los cursos académicos. Contiene el año académico (si es el curso 2018-2019 se pone el primero de los años, en este caso es 2018), la fecha de inicio y fin del curso académico, flag que indica si es el curso actual y las fechas de inicio y fin de la Constitución de las Juntas Electorales, de admisión de candidaturas a los Consejos Escolares, de las campañas electorales a los Consejos Escolares y de las elecciones a los Consejos Escolares.

**TLDATCENCURSO:** Tabla que almacena los datos de los centros por curso. Contiene el año académico, el ID del centro, el ID del tipo de jornada, la fecha de inicio y fin del periodo de horario regular, un flag que indica si el centro está autorizado a desarrollar planes de compensación educativa y un flag que indica si los profesores pueden definir el horario.

**TLDATESPEVAMAT:** Tabla de datos sobre la evaluación de las matrículas. Contiene las observaciones que el tutor hace sobre la evaluación, el número de convocatoria asociada a los datos y la matrícula del alumno.

**TLDATESPEVAMATMAT:** Tabla de datos sobre la evaluación de las materias de una matrícula. Contiene la convocatoria de evaluación a la que está asociada los datos, el ID de materia y las observaciones que hace el tutor sobre el alumno.

**TLDATEXPHISTALU:** Tabla de los datos del expediente histórico del alumno. Contiene la matrícula del alumno, el centro, datos médicos y psicológicos, si convalida por estudios superiores de música o danza o por ser deportista de alto rendimiento y las observaciones pertinentes.

**TLDATGENSOLAYU:** Tabla de los datos genéricos de la solicitud de ayuda. Contiene el ID de solicitud, la convocatoria, el ID de alumno, el resultado de la convocatoria de ayuda, el motivo de exclusión de ayuda (en caso de haber exclusión), el estado de la solicitud y si presenta documentación de discapacidad.

**TLDATOSCEN:** Tabla de los datos del centro. Contiene el ID de centro, el ID de municipio y provincia, las fechas de inicio y fin de vigencia, una descripción del centro, el código postal, el teléfono, fax, correo electrónico y el CIF del centro.

**TLDEPACTEXT:** Tabla de departamentos de actividades extraescolares. Contiene el ID de la actividad extraescolar y el ID del departamento de la actividad extraescolar.

**TLDEPARTAMENTOS:** Tabla de departamentos. Contiene su ID y una descripción.

**TLDEPENDENCIAS:** Tabla de dependencias o huecos (aulas). Contiene el ID del aula, su descripción larga y corta, el tipo de dependencia, el ID de la planta del edificio, un flag indicando si está activa esa dependencia, el número de puestos que tiene (capacidad), el ancho, largo, alto y superficie del aula y el número máximo que puede ser reservada por empleado.

**TLDISCALUM:** Tabla de discapacidades de alumnos. Contiene el ID de discapacidad, una descripción larga y otra corta, el ID de alumno y si tiene autonomía.

**TLDISCAP:** Tabla de discapacidades. Contiene el ID del tipo de discapacidad y una descripción de la misma.

**TLEDIFICIOS:** Tabla de edificios. Contiene el ID del edificio, una descripción larga y otra corta del mismo, el ID del centro al que pertenece, un flag indicando si es el edificio principal, la dirección, la localidad, el teléfono, fax y correo electrónico.

**TLEDIPLA:** Tabla de plantas de los edificios. Contiene el ID del edificio al que pertenece esa planta y una descripción.

**TLEMPLEADOS:** Tabla de los empleados en los centros. Contiene el ID del empleado, el ID del municipio y de la provincia, el DNI del empleado, el nombre y dos apellidos, el sexo del empleado, la fecha de nacimiento, la calle donde vive, el número, si tiene escalera, el código postal, el teléfono, el número de la seguridad social, el centro del puesto de trabajo activo, el puesto de trabajo y número de años de servicio.

**TLESPECIALIDADES:** Tabla de las diferentes especialidades de los empleados (arqueología, alfarería, técnicas gráficas, matemáticas...). Contiene el código de especialidad y una descripción larga y otra corta de la misma.

**TLESTADOMAT:** Tabla del estado de la materia. Contiene el ID de la materia y la descripción (matriculada, pendiente, convalidada...).

**TLESCIVIL:** Tabla del estado civil. Contiene su ID y dos descripciones, una larga y otra corta.

**TLESTGENMATR:** Tabla del estado genérico de la matrícula. Contiene el ID de estado, un código de resultado y la descripción de ese estado (renuncia, trasladada, obtiene título...).

**TLETAPAS:** Tabla que representa a las etapas. Contiene su ID, su descripción (1º Primaria, 2º Primaria...), si es de estudios obligatorios y su año de inicio y de fin.

**TLETAPORCEN:** Tabla de las etapas que tiene cada centro. Contiene el ID de etapa, el ID de centro y la fecha de inicio y fin de vigencia.

**TLEXPCENTROS:** Tabla de los expedientes de cada centro. Contiene el ID de alumno, el tipo de expediente, el ID del centro, el año académico, el número de expediente, flag que indica si el expediente ha sido trasladado, si el expediente procede de un traslado, la fecha de traslado, el centro de origen y de destino del expediente y la localidad de origen y destino del expediente.

**TLEXPEDIENTES:** Tabla que guarda todos los expedientes. Contiene el ID de alumno, el año académico, el número de matrículas del alumno en ese tipo de expediente y el número de repeticiones de cursos del alumno en ese expediente.

**TLFALASIALU:** Tabla donde se registran las faltas de asistencia del alumno. Contiene el ID del tramo horario de la ausencia, el ID de la matrícula, la fecha de la falta de asistencia, el tipo de falta (justificada, injustificada o retraso) y un flag indicando si falta todo el día completo.



Merece la pena destacar que en RACIMA, si un profesor falta a clase, esa falta se llama ausencia, no falta de asistencia.

**TLGRUTIPCONNEG:** Tabla que almacena el tipo de conductas contrarias en grupo. Contiene el ID del tipo de conducta contraria grupal y su descripción.

**TLGRUTIPCORCONNEG:** Tabla de correcciones de las conductas contrarias. Contiene el ID del tipo de corrección y su descripción.

**TLHORARIOCEN:** Tabla en la que se guardan los horarios del centro. Contiene el ID de centro, si es IES o escuela, el año, un flag de si la actividad es de docencia directa, el día de la semana, la hora de inicio y fin de la actividad, número de minutos de la actividad, DNI del tutor de la unidad, la actividad realizada, la materia y un flag que indica si la actividad pertenece al horario.

**TLHORARIOSR:** Tabla de horarios del profesor. Esta tabla es muy importante dado que es la pieza central de los profesores. Todo lo relacionado con impartir clases, profesores y horarios de alumnos vienen del horario del profesor. Contiene el ID del horario, el ID de actividad, el ID del empleado, el año académico, la fecha de inicio y fin de cada tramo, la hora de inicio y fin de cada tramo y el número de minutos computados en el horario del profesor.

**TLINFALUENSEG:** Tabla de informes de alumnos en seguimiento. Contiene el ID de alumno, el ID del tipo de informe, el ID del empleado que hace el informe y una descripción del tipo de informe.

**TLJORBAS:** Tabla de horas lectivas o complementarias por día. Contiene el ID de centro, el ID del tipo de personal, el año académico y las horas lectivas por cada día (de lunes a domingo).

**TLJUSMOT:** Tabla de justificación de ausencias. Contiene el ID de la justificación y el motivo de la ausencia.

**TLMATABSESC:** Tabla de matrículas con absentismo escolar. Contiene el ID de la matrícula del alumno absentista y observaciones indicativas del motivo de absentismo.

**TLMATALU:** Tabla de matrículas de alumnos. Esta tabla es de gran importancia, dado que el estudio se basa en esta tabla casi siempre para detectar los alumnos con fracaso y sin fracaso. Contiene el ID de matrícula, el ID de alumno, el ID del centro en el que está matriculado, el año académico, el período de matriculación, el tipo de matrícula, el código del municipio, la localidad, el estado de la matrícula (obtiene título, no obtiene título, no promociona...), el resultado de la matrícula (2 significa que obtiene el título, 3 que no lo obtiene), el grupo al que pertenece el alumno y en qué curso está matriculado (1ºESO, 2º Primaria...).

**TLMATERIASCURSO:** Tabla de las materias por curso. Contiene una descripción por cada materia, indica si agrupa a otras materias, si está vigente, si se imparte de forma individual y si precisa de autorización para impartirla.

**TLMATMATRIALU:** Tabla de materias por matrícula de alumno. Contiene el ID de la matrícula, el estado de la materia matriculada, flag que indica si está aprobada, ID de la materia del centro, ID de la calificación y la fecha de aprobado (en caso de estar aprobada).

**TLMATUNIACTEXT:** Tabla de matriculación en actividades extraescolares. Contiene el ID de la actividad extraescolar, el ID de matrícula, flag que indica si el alumno convocado está inscrito, flag que indica si el alumno presenta autorización y flag que indica si el alumno falta a la actividad extraescolar.

**TLMEDALUABS:** Tabla de medidas para un alumno absentista. Contiene el ID del alumno absentista, la fecha de la medida y el nombre de la persona que realiza la actuación sobre un expediente de absentismo.

**TLMEDPROCONNEG:** Tabla de medidas provisionales de las conductas negativas o contrarias. Contiene una descripción de cada medida provisional larga, otra corta y el código de la medida.

**TLMOTAUS:** Tabla de motivos de ausencias. Contiene una descripción larga para cada motivo, la fecha de inicio y fin de vigencia del motivo, si el motivo es por huelga y si la ausencia está injustificada.

**TLMUNICIPIOS:** Tabla de municipios. Contiene el ID de la provincia, una descripción corta y una descripción larga.

**TLMUNINF:** Tabla de municipios de influencia. En esta tabla están los centros conocidos como RAC. Contiene el ID de la provincia, el ID de municipio y el ID del centro al que pertenecen.

**TLNIVCEN:** Tabla que almacena los distintos niveles de los centros. Contiene el ID del nivel educativo, el ID de centro, el año desde el cual se imparte ése nivel educativo y el año en el que se termina de impartir dicho nivel.

**TLNIVEDU:** Tabla de niveles educativos. Contiene una descripción breve de cada nivel educativo.

**TLNOTASMEDIAS:** Tabla de notas medias. Contiene el ID del tipo de nota media, el ID de alumno, la nota media, flag que indica si tiene matrícula de honor y el ID de calificación.

**TLNOTEVA:** Tabla de las notas de la evaluación. Contiene el ID de la convocatoria, el ID de la matrícula, el ID de calificación y la nota obtenida.

**TLOBSEVAL:** Tabla de observaciones en la evaluación. Contiene el ID de matrícula, el ID de la convocatoria y la observación sobre la evaluación de la matrícula.

**TLOTTRDATMATEST:** Tabla de otros datos relevantes de matrículas asociados a estadísticas. Contiene el ID de matrícula, un flag que indica si está trabajando por cuenta propia o ajena, un flag que indica si vive con 1 sólo progenitor, flag que indica si vive en un centro de acogida, flag que indica si vive en infravivienda o en lugar desconocido, flag que indica si posee discapacidad, flag que indica si presenta

necesidades de apoyo educativo y un flag que indica si en el curso actual está trabajando.

**TLPAIS:** Tabla que almacena todos los países. Contiene el código de cada país, una descripción corta, otra larga de ese código y si ese código es España.

**TLPARAUS:** Tabla de los partes de ausentismo. Contiene el ID del centro, el año académico, el mes, el orden del parte, el año natural, la fecha de remisión y algunas observaciones.

**TLPARCENBIL:** Tabla de partes del centro bilingüe. Contiene el ID del proyecto del centro bilingüe, el ID de profesor, los estudios realizados por el participante, la comprensión oral, la comprensión escrita, la expresión oral, la expresión escrita, la comprensión auditiva, la comprensión de lectura, flag de si solicita actividades de perfeccionamiento del idioma en la localidad y otro flag de si solicita actividades de perfeccionamiento del idioma en el extranjero.

**TLPERSONAS:** Tabla de personas. Aquí está tanto el personal docente como el no docente. Contiene el ID de la persona, el tipo de identificación (DNI o pasaporte), el nombre, el primer y segundo apellido y la fecha de nacimiento.

**TLPROVINCIAS:** Tabla de provincias. Contiene el ID de la provincia, el ID del país y una descripción larga de la provincia.

**TLPTOTRAEMP:** Tabla de puestos de trabajo de los empleados. Contiene el ID del empleado, la fecha de toma de posesión, la fecha de cese de posesión, el ID del centro, el ID del puesto, el número de años de permanencia en el centro, la fecha de toma de posesión real y la fecha de cese real.

**TLPUUESTOS:** Tabla de puestos de trabajo. Contiene el ID del puesto, una descripción corta del puesto, una descripción larga del puesto, la fecha de inicio de vigencia y la fecha de fin de vigencia.

**TLREFLINALUEXT:** Tabla de refuerzo lingüístico a alumnos extranjeros. Contiene el ID de la matrícula del alumno extranjero, el ID del modo de refuerzo, el número de horas a la semana y el ID del organizador.

**TLRESACTEXT:** Tabla de los responsables de las actividades extraescolares. Contiene el ID de la actividad extraescolar y el ID del responsable.

**TLRESALPRUPSI:** Tabla de resultados de las pruebas psicopedagógicas. Contiene el ID del alumno, el resultado, el ID de la subescala de las pruebas y el ID de puntuación.

**TLRESULTADO:** Tabla de los resultados de las matrículas (1 es cancelada, 2 superada, 3 no superada). Contiene el código del resultado de la matrícula y una descripción larga del resultado.

**TLISISCAL:** Tabla del sistema de calificación. Contiene una descripción larga de cada sistema, una descripción corta, flag que indica si usa escala numérica, flag que indica si es un sistema cerrado y la nota del umbral del aprobado para cada sistema numérico.

**TLAYULIBMATEESC:** Tabla de solicitudes de ayudas con los libros de las matrículas escolarizadas. Contiene el ID de la ayuda al estudio, el ID de matrícula y un flag que indica si presenta una grave necesidad sociofamiliar.

**TLTIPACTEXT:** Tabla de tipos de actividades extraescolares. Contiene una descripción larga para cada tipo de actividad, una corta para cada tipo y el código del tipo de actividad extraescolar.

**TLTIPCONNEG:** Tabla de tipos de conducta negativa o contraria. Contiene una descripción corta, otra larga y otra completa para cada tipo de conducta contraria, además de un flag que indica si permite apertura de expediente.

**TLTIPCORCONNEG:** Tabla de tipo de corrección de conductas contrarias o negativas. Contiene una descripción larga para cada tipo, un flag que indica si conlleva amonestación y otro flag que indica si conlleva expulsión.

**TLTIPEXPCONNEG:** Tabla de tipos de expediente de conductas negativas o contrarias. Contiene una descripción corta para cada tipo de expediente, otra larga para lo mismo y el código del tipo de expediente.

**TLTIPJUS:** Tabla de los tipos de justificación. Contiene una descripción de cada tipo y su código.

**TLTIPCEN:** Tabla de los tipos de centros. Contiene una descripción larga y otra corta de cada tipo de centro.

**TLTIPPRUPSI:** Tabla de tipos de pruebas psicopedagógicas. Contiene el ID de la categoría de las pruebas psicopedagógicas, un código del tipo de prueba estandarizada y una descripción de cada tipo de prueba psicopedagógica.

**TLTRAMOSHOR:** Tabla con los tramos horarios. Contiene el año académico del tramo, el ID del centro, una descripción larga de la hora del centro, la hora de inicio del tramo, la hora de fin del tramo y el número de minutos que computará en el horario de los profesores.

**TLTUTORES:** Tabla con los tutores de los alumnos. Contiene los nombres del padre y de la madre, sus apellidos, sus teléfonos, sus correos electrónicos, su localidad, su provincia, su país y el domicilio donde viven.

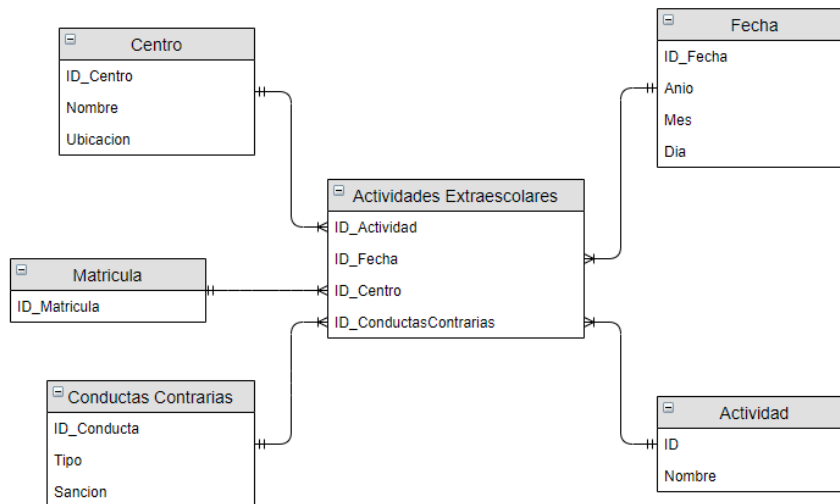
Esas tablas son las más interesantes para el estudio. Tras una reunión con el cliente, se comentó que varias de esas tablas estaban vacías porque no las utilizaban en RACIMA. Con el cliente se estuvo estudiando cuáles de esas tablas podrían ser útiles y cuáles no. El motivo de la elección de estas tablas reside en las ideas y percepciones que se tiene sobre el fracaso escolar que se han puesto en común con el cliente, debatiendo las tablas que pueden arrojar algo de ayuda en el estudio y las que no aportan nada. Tras este intercambio de ideas, se seleccionaron las siguientes tablas:

- **TLALUCONCON**
- **TLALUMNOS**
- **TLAREPRUDIA**
- **TLASIGIMPHORCLA**
- **TLAUSENCIAS**
- **TLCALIFICAIONES**
- **TLCENADS**
- **TLCENADSIES**

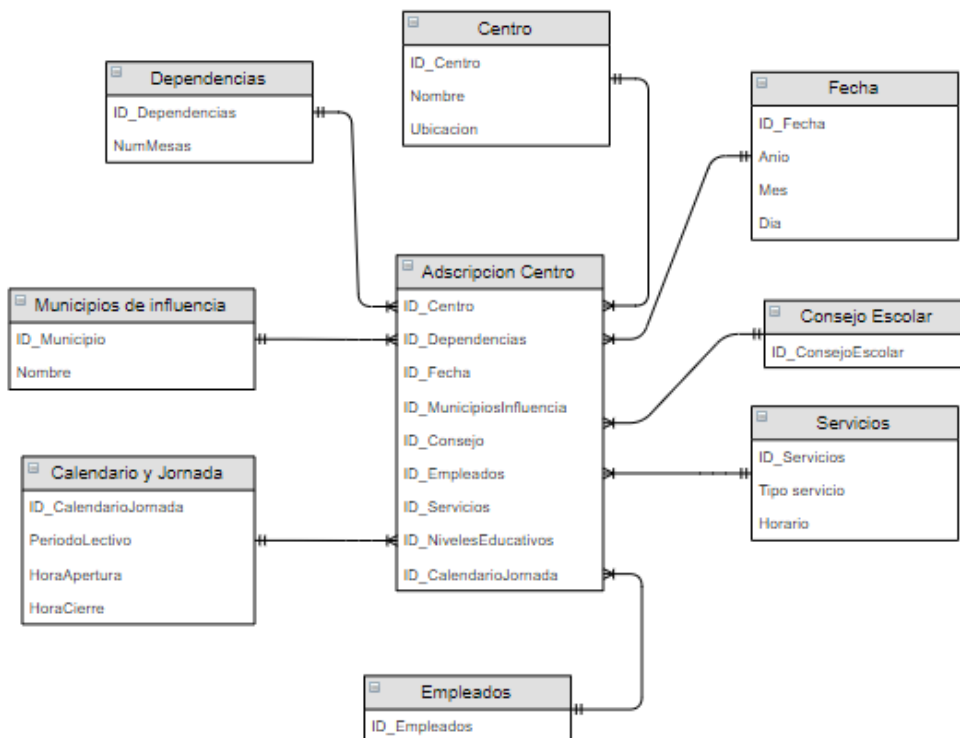
- **TLCENTROS**
- **TLCONNEGALU**
- **TLCURSOACA**
- **TLEMPLEADOS**
- **TLESTADOMAT**
- **TLESTGENMATR**
- **TLFALASIALU**
- **TLHORARIOSR**
- **TLJUSMOT**
- **TLMATALU**
- **TLMATMATRIALU**
- **TLMOTAUS**
- **TLMUNICIPIOS**
- **TLPAIS**
- **TLPERSONAS**
- **TLPROVINCIAS**
- **TLRESULTADO**
- **TLTIPCONNEG**
- **TLTRAMOSHOR**

### 3. Capítulo 2: diseño de hechos y dimensiones

Tras el estudio de las vistas que se proporcionaron, se adelantó trabajo creando posibles “estrellas”, es decir, hechos con sus dimensiones. Estas tablas fueron diseñadas cuando el único recurso que se disponía eran las vistas en formato papel que fueron facilitadas por el cliente. En las figuras 3.1 a 3.10 se muestran las estrellas que inicialmente fueron diseñadas:

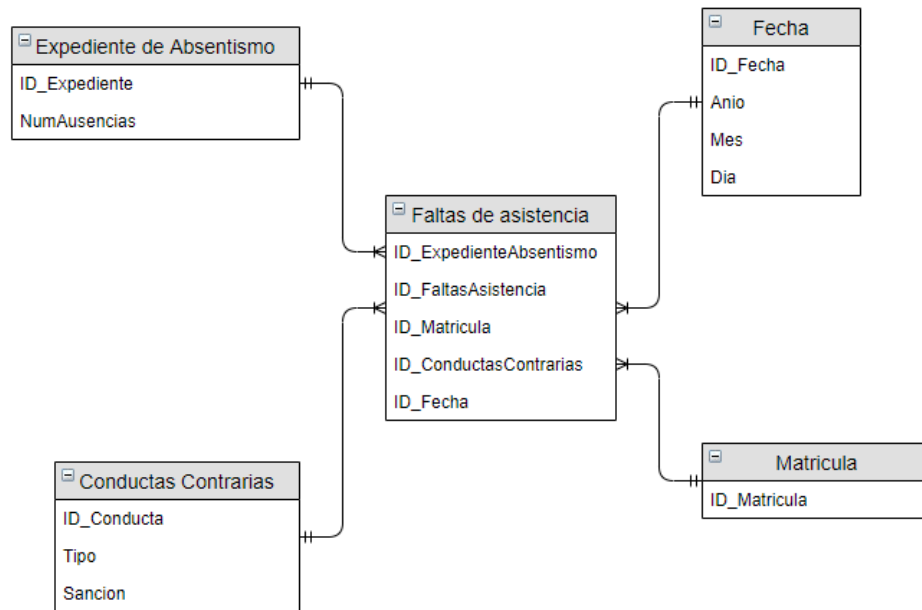


**Figura 3.1.** Estrella del hecho Actividades extraescolares

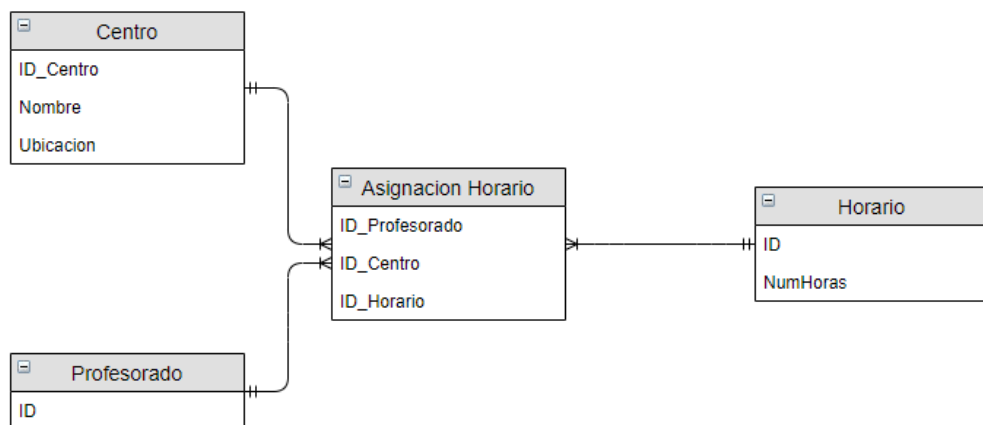


**Figura 3.2.** Estrella del hecho Adscripciones de centros





**Figura 3.3.** Estrella del hecho Faltas de asistencia



**Figura 3.3.** Estrella del hecho Asignaciones de horario a un profesor

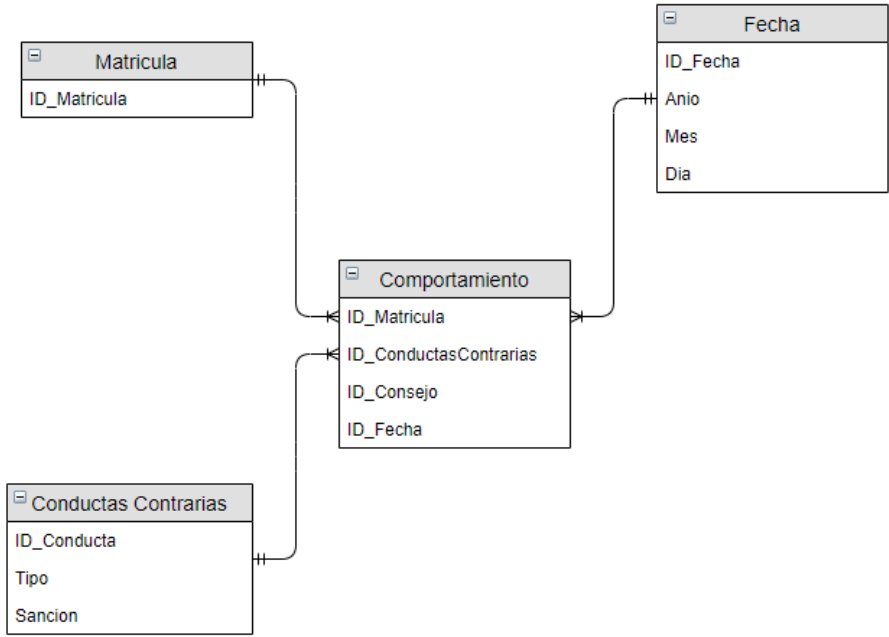


Figura 3.4. Estrella del hecho comportamiento

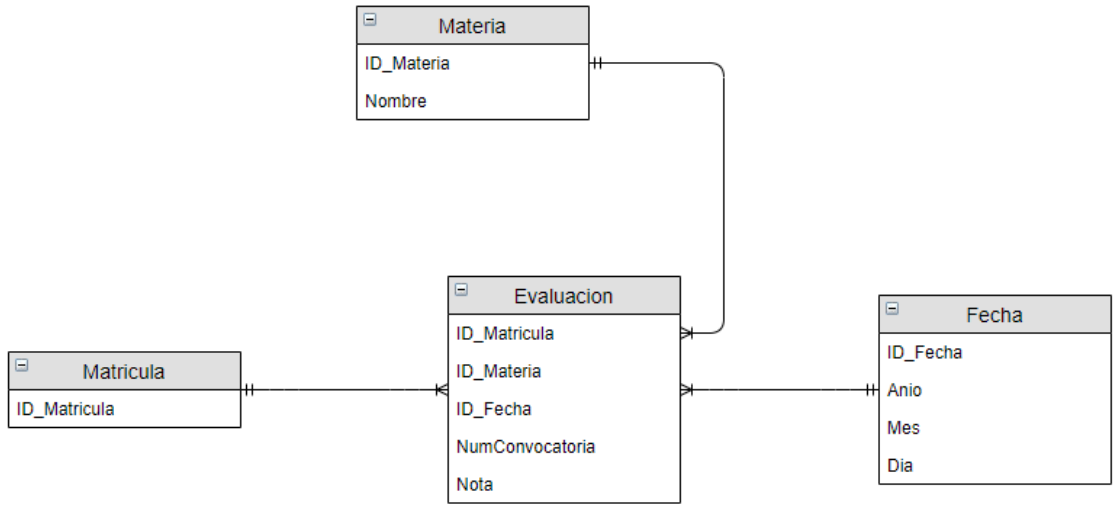
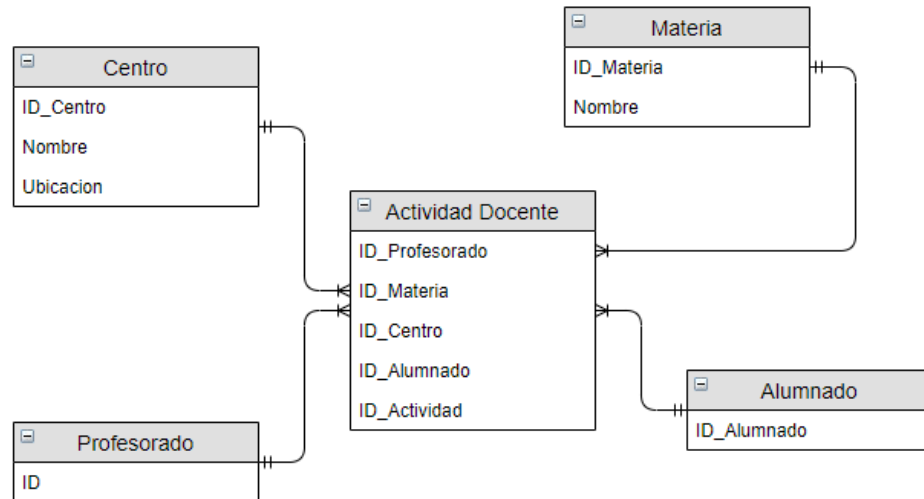
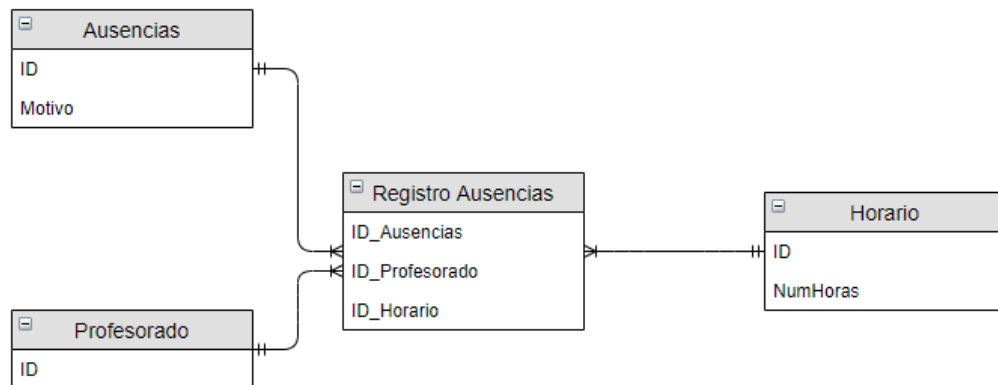


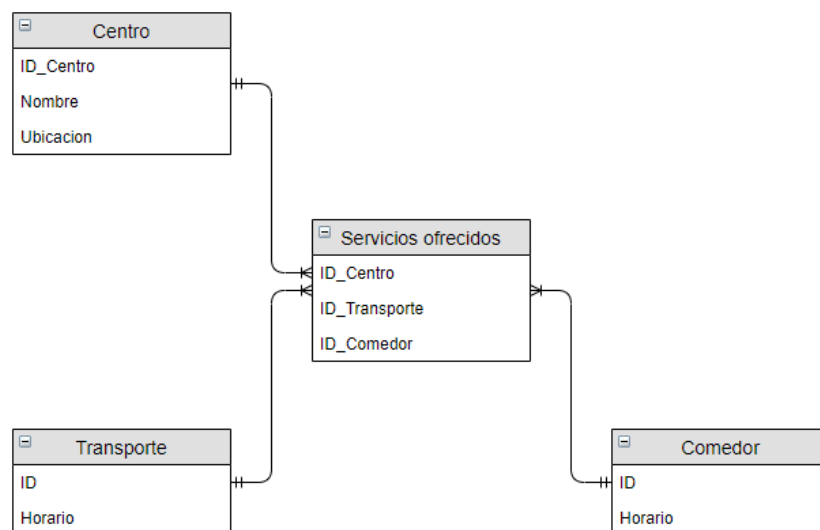
Figura 3.5. Estrella del hecho evaluaciones



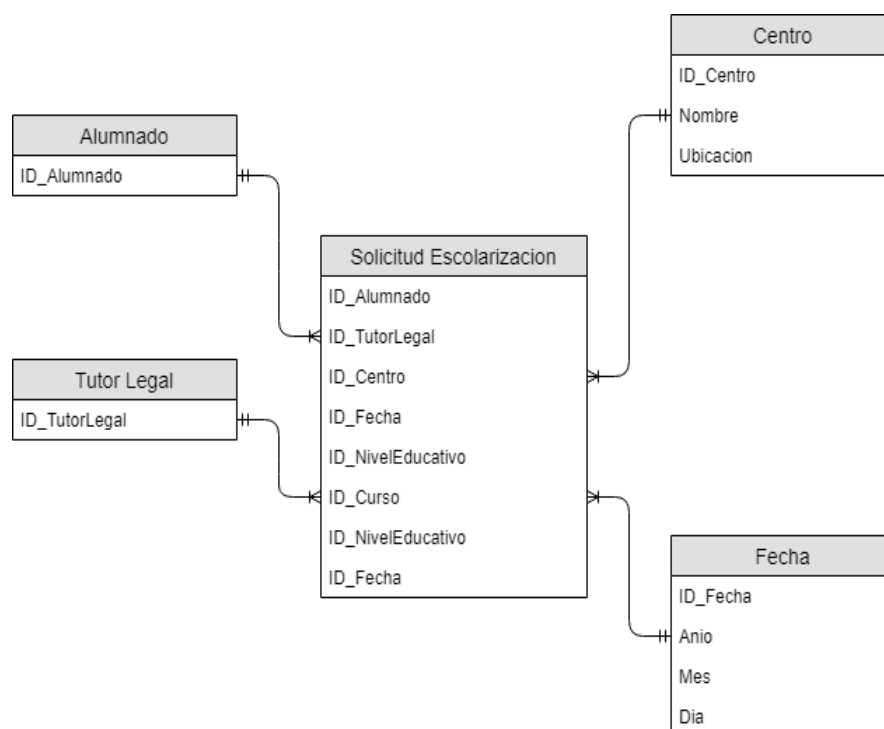
**Figura 3.6.** Estrella del hecho Actividad docente



**Figura 3.7.** Estrella del hecho Registro de ausencias



**Figura 3.8.** Estrella del hecho Servicios ofrecidos



**Figura 3.9.** Estrella del hecho Solicitud de escolarización

## 4. Capítulo 3: comparaciones entre alumnos con fracaso y sin fracaso

En este capítulo, una vez establecidas las tablas que van a ser de utilidad para el proyecto, se procedió a analizar a los alumnos con fracaso y observar similitudes entre ellos y compararlos con los que no han obtenido fracaso.

Debido a que es la primera vez que se hace ETL sobre la base de datos de RACIMA y que no se tiene un trabajo previo sobre el que basarse, este estudio se ha acordado con el cliente que se basará principalmente en el concepto percibido de fracaso escolar, orientando la búsqueda hacia lo más evidente del fracaso escolar.

El primer paso fue remarcar bien la condición que separa a los alumnos con fracaso escolar de los que no, que es la obtención del título de la E.S.O. una vez cursado el último curso. A partir de esa definición se crearon tres vistas para tener a los alumnos bien separados. Una vista contenía a los alumnos de 4º de la E.S.O., la segunda vista contenía a los alumnos de 4º de la E.S.O. que tienen título (es decir, los que no fracasan) y, por último, la vista de los alumnos que no han obtenido el título de la E.S.O. (los que tienen fracaso escolar). Para hacer esta última vista no bastaba sólo con hacer un select de los alumnos que tenían en la matrícula que no habían obtenido título (como sí se ha hecho para los alumnos con título) porque se puede dar el caso de que un alumno no obtenga título en un año concreto y al siguiente sí lo consiga. En este caso particular, dicho alumno aparecería en ambas vistas, la de alumnos con fracaso y la de alumnos sin fracaso. Para evitar esto, se ha hecho uso de la sentencia **minus**.

Estas han sido las consultas:

- **Vista de alumnos de 4º de la E.S.O.**

```
create view alumnos_cuarto as
(select distinct mua.x_alumno
from tlmatalu mua, tlofematrgen omg
where OMG.X_OFERTAMATRIG=MUA.X_OFERTAMATRIG
and (omg.t_abreviatura like '%4ºESO%' or omg.t_abreviatura like '%4º E.S.O.%') and
(mua.c_anno between 2011 and 2017));
```

Con esta consulta, se obtienen todos los alumnos en cuya oferta de matrícula esté marcado 4º E.S.O y además sólo se obtienen los alumnos referentes a los años comprendidos entre 2011 y 2017, debido a que RACIMA empezó a tener datos consistentes y completos a partir de ese año a pesar de llevar más tiempo activa y hasta el 2017 porque no se tienen los datos del año 2018 dado que se acababa de cerrar ese año escolar y faltaban datos.

- **Vista de alumnos de 4º de la E.S.O. sin fracaso escolar**

```
create view consigue_titulo as
(select distinct mua.x_alumno
from tlmatalu mua, tlofematrgen omg
where OMG.X_OFERTAMATRIG=MUA.X_OFERTAMATRIG
and (omg.t_abreviatura like '%4ºESO%' or omg.t_abreviatura like '%4º
E.S.O.%') and (mua.c_resultado = 2) and (mua.c_anno between 2011 and
2017));
```

Esta consulta es muy similar a la anterior con la única novedad de que se ha añadido que el campo resultado sea 2 (Figura 4.1). Esto es porque si en ese campo hay un 2, significa que esa matrícula obtuvo el título de 4º de la E.S.O., si tiene un 3 significa que no obtuvo el título y si tiene un 1 significa que la matrícula está cancelada.

	⚡ C_RESULTADO	⚡ D_RESULTADO	⚡ T_ABREV	⚡ L_PROMOC...
1	2	Superada	SUP	S
2	3	No superada	NSUP	N
3	1	Cancelada	CAN	N

**Figura 4.1.** Tabla de resultados de una matrícula

- **Vista de alumnos de 4º de la E.S.O. con fracaso escolar**

```
create view no_obtiene_titulo as
(select distinct mua.x_alumno
from tlmatalu mua, tlofematrgen omg
where OMG.X_OFERTAMATRIG=MUA.X_OFERTAMATRIG
and (omg.t_abreviatura like '%4ºESO%' or omg.t_abreviatura like '%4º
E.S.O.%') and (mua.c_resultado = 3) and (mua.c_anno between 2011 and 2017)
minus
select distinct mua.x_alumno
from tlmatalu mua, tlofematrgen omg
where OMG.X_OFERTAMATRIG=MUA.X_OFERTAMATRIG
and (omg.t_abreviatura like '%4ºESO%' or omg.t_abreviatura like '%4º
E.S.O.%') and (mua.c_resultado = 2) and (mua.c_anno between 2011 and
2017));
```

Tras tener las vistas, se procedió a comparar a los alumnos que tenían fracaso con los que no siguiendo una serie de pautas que fueron acordadas con el cliente. Se pusieron en común los aspectos que podría ser interesantes estudiar de los alumnos.

Primeramente, destacar que este estudio está basado en porcentajes, dado que el número de alumnos con título era superior al de alumnos sin él. Se incluyen las consultas SQL que se han utilizado para obtener los resultados. Únicamente se muestran las que se utilizaron para obtener datos de los alumnos sin título, dado que las consultas de los alumnos con título son idénticas salvo porque se emplea la vista **consigue\_titulo** en vez de **no\_obtiene\_titulo**.

Si se incluyen ambas consultas, es porque se tuvo que hacer variaciones en alguna de ellas para obtener los datos.

Alumnos de 4º de la E.S.O.: 15240

Alumnos de 4º de la E.S.O sin título: 1291

Alumnos de 4º de la E.S.O con título: 13759

- **Alumnos que repitan curso**

Se compararon a los alumnos para ver en qué grupo había alumnos que repetían más veces curso, y se obtuvo que un 86,67% de los alumnos sin título repiten algún curso, mientras que sólo un 29,46% de los alumnos con título repiten alguna vez. Esto indica que si un alumno repite curso, puede ser indicativo de fracaso escolar.

A continuación, se muestra la consulta para obtener a los alumnos sin título que repiten:

```
select count(distinct x_alumno)
from tlmatalu
where x_alumno in (select x_alumno from no_obtiene_titulo) and l_repite='S';
```

- **Alumnos con conductas contrarias**

Tras comparar a los alumnos para ver en qué grupo había más alumnado con conductas contrarias, se observó que en el grupo de los alumnos sin título había un 51,51% de alumnos que tenían conductas contrarias, mientras que en el grupo de los alumnos con título, había un 37,51% de ellos con conductas contrarias. Esto significa que un alumno con conductas contrarias, tiene bastantes posibilidades de acabar con fracaso escolar.

A continuación se observa la consulta para obtener a los alumnos sin título que tienen conductas contrarias:

```
select count (*)
from (select distinct tit.x_alumno, count(*) as total_con_con
      from no_obtiene_titulo tit join
      tlmatalu alu on tit.x_alumno=alu.x_alumno
      join tlaluconcon alucon on alu.x_matricula=alucon.x_matricula group by
      tit.x_alumno);
```

- **Alumnos con faltas de asistencia**

En el estudio de las faltas de asistencia, ambas partes presentan un alto porcentaje de faltas de asistencia, pero los alumnos sin título tienen un porcentaje mayor (un 89,70% frente a un 82,92%).



Para este estudio, se decidió diferenciar el tipo de faltas de asistencia, dado que las faltas de asistencia pueden ser de 3 tipos: justificada, injustificada o retraso.

Se obtuvo el total de faltas de asistencia de ambos grupos (470.707 faltas entre los alumnos sin título y 2.208.012 entre los alumnos con título) y luego se clasificaron las faltas según su tipo.

En las faltas injustificadas, hay una mayoría entre los alumnos sin título (83,27% frente a un 72,52%), en las faltas justificadas ocurre lo contrario, que los alumnos con título tienen más faltas justificadas (22,46% frente a un 12,48% de los alumnos sin título) y los alumnos con título han tenido más retrasos que los que no tienen título, pero estos porcentajes son muy similares (4,25% de alumnos sin título frente a un 5,02% de alumnos con título). Esto significa que si un alumno tiene faltas de asistencia, puede ser indicativo de fracaso escolar, pero si además la falta es injustificada, es más probable que el alumno fracase.

La consulta para obtener a los alumnos sin título que tienen Faltas de asistencia se muestra a continuación:

```
select count (*) as faltas
from no_obtiene_titulo sitit join
tlfalatalu alu on sitit.x_alumno=alu.x_alumno join
tlfalasialu falasi on alu.x_matricula=falasi.x_matricula;
```

Para obtener el tipo de falta (justificada, injustificada o retraso) basta con añadir la cláusula where así: `where falasi.c_tipfal=`, donde `c_tipfal` puede ser R, J o I (Retraso, Justificada o Injustificada).

- **Media de las notas de los alumnos**

Para obtener la media de sus expedientes, se realizó una subconsulta que obtenía para cada alumno con fracaso la media de las notas de cada materia y luego se obtuvo la media de todas las medias de los alumnos con fracaso (para los alumnos con título se hizo lo mismo). La media de los alumnos sin título es de un 5,11 y la de los alumnos con título es de un 6,85. Esto indica que si un alumno suspende muchas veces, puede derivar en fracaso.

A continuación se muestra la consulta para obtener las medias de los alumnos sin título:

```
select avg (media) from (select notit.x_alumno, avg(cal.n_numero) as media  
from no_obtiene_titulo notit join tlmatalu alu on notit.x_alumno=alu.x_alumno  
join tlmattmatialu mat on mat.x_matricula=alu.x_matricula join tlcalfificaciones  
cal on mat.x_califica=cal.x_califica group by notit.x_alumno);
```

- **Alumnos españoles y extranjeros**

Dado que pueden darse problemas de aprendizaje debido a los diferentes idiomas, se decidió analizar si había más fracaso escolar entre los alumnos de nacionalidad española o entre los alumnos de nacionalidad extranjera. Para ello, se dividió a los alumnos de 4º de la E.S.O. en 2 grupos, los que tienen nacionalidad española y los que no. Luego se comprobó el tanto por ciento de alumnos españoles y extranjeros con fracaso, y resultó que un 6,10% del alumnado español fracasa y un 24,04% del alumnado extranjero fracasa. De ahí se puede ver cómo los alumnos extranjeros son más propensos a fracasar.

La consulta para obtener los alumnos españoles se puede ver a continuación:

```
select count (distinct(al.x_alumno)) from alumnos_cuarto al join tlalumnos alu  
on al.x_alumno=alu.x_alumno where alu.c_pais_nacer=034;
```

Para obtener a los alumnos no españoles, basta con modificar el where así:

```
where alu.c_pais_nacer!=034
```

Consulta para obtener a los alumnos españoles con fracaso:

```
select count (*) from (select distinct(al.x_alumno) from alumnos_cuarto al join  
no_obtiene_titulo notit on al.x_alumno=notit.x_alumno join tlalumnos alu on  
al.x_alumno=alu.x_alumno where alu.c_pais_nacer=034
```

minus

```
select distinct(al.x_alumno) from alumnos_cuarto al join consigue_titulo notit on  
al.x_alumno=notit.x_alumno join tlalumnos alu on al.x_alumno=alu.x_alumno  
where alu.c_pais_nacer=034);
```

Se ha programado de esta manera para evitar el caso en el que un alumno no obtenga el título un año, pero al siguiente sí, porque en ese caso aparecería dos veces y sólo interesan los alumnos que no hayan obtenido el título de ninguna manera. La idea es la misma que la utilizada al principio para obtener al alumnado con fracaso escolar.

- **Hombres y mujeres**

La diferencia entre hombres y mujeres a la hora de ver qué grupo tiene más fracaso ha dado resultados con poca diferencia, siendo un 10,04% del total de hombres los que fracasan y un 7,05% de mujeres que fracasan.

La consulta para obtener a los alumnos hombres con fracaso es la siguiente:

```
select count (*) from (select distinct(al.x_alumno) from alumnos_cuarto al join
tlalumnos a on al.x_alumno=a.x_alumno join no_obtiene_titulo notit on
al.x_alumno=notit.x_alumno where l_sexo='H'
minus
select distinct(al.x_alumno) from alumnos_cuarto al join tlalumnos a on
al.x_alumno=a.x_alumno join consigue_titulo notit on
al.x_alumno=notit.x_alumno where l_sexo='H');
```

- **Tipo de centro (privado o público)**

Con el objetivo de comprobar en qué tipo de centros se da más fracaso, se obtuvo el número de alumnos con fracaso en cada tipo de centro. El resultado fue que el porcentaje de alumnos de 4º de la E.S.O. que fracasan en centros públicos es del 9% mientras que el porcentaje de alumnos con fracaso en centros privados es del 7,65%.

A continuación se muestra la consulta para obtener a los alumnos con fracaso escolar en centros públicos (para centros privados basta con cambiar el valor de cen.x\_tipo por 2):

```
select count(distinct (al.x_alumno)) from alumnos_cuarto al join
no_obtiene_titulo notit on al.x_alumno=notit.x_alumno join tlmatalu alu on
notit.x_alumno=alu.x_alumno join tlcentros cen on alu.x_centro=cen.x_centro
where cen.x_tipo=1;
```

- **Municipio del centro y de los alumnos**

Para comprobar si vivir en un municipio diferente al del centro influye en el fracaso, se decidió comprobar cuántos de los alumnos con fracaso escolar viven en municipios diferentes a su lugar de residencia. El resultado ha sido que de los alumnos con fracaso, un 34,08% viven en municipios diferentes al centro, mientras que entre el alumnado sin fracaso, sólo el 3,64% viven en municipios distintos al del centro. Se observa mucha diferencia entre ambos porcentajes, lo que indica que vivir en un municipio diferente es significativo para el fracaso escolar.

La consulta para obtener los alumnos con fracaso que viven en municipios distintos al del centro se puede observar a continuación:

```
select count(distinct (notit.x_alumno)) from no_obtiene_titulo notit join tlmatalu  
alu on notit.x_alumno=alu.x_alumno join tlalumnos alum on  
notit.x_alumno=alum.x_alumno join tldatoscen dat on  
alu.x_centro=dat.x_centro where alum.c_municipio!=dat.c_municipio;
```

## 5. Capítulo 4: implementación de los hechos y dimensiones

Una vez acabado el proceso de comparación entre alumnos con fracaso y sin él, se ha procedido a implementar las tablas de dimensiones y de hechos que constituirán el Data Warehouse. Como se ha mencionado previamente, para este punto se utilizará la herramienta Pentaho.

Observando los resultados del estudio de los alumnos y debido a que queremos ver los resultados por alumnos, se ha optado por crear una tabla de hechos H\_ALUMNOS que contendrá las dimensiones que se explican a continuación:

- D\_CURSO: Dimensión que contiene los diferentes cursos registrados en la base de datos. Consta de dos campos, el primero es un año (el primero de los años que comprende el curso, es decir, en el curso 2017-2018 se indicará el año 2017) y el segundo son los años que abarca ese curso.
- D\_REPETIDOR: Dimensión que contiene dos campos. Uno contiene una “S” o una “N” y el segundo es un descriptivo de esos campos. Esta dimensión indica si un alumno es repetidor o no.
- D\_CONDUCTAS\_CONTRARIAS: Dimensión que almacena si hay registros de conductas contrarias de un alumno. El diseño es idéntico al de la dimensión D\_REPETIDOR.
- D\_FALTAS\_ASISTENCIA: Dimensión que almacena dos datos. El primero de ellos es un número que indica a qué grupo de faltas de asistencia pertenece un alumno. El segundo dato es un descriptivo del grupo (por ejemplo, el grupo 1 son los que tienen entre 1 y 99 faltas de asistencia).
- D\_NACIONALIDAD: Dimensión con la misma estructura que la dimensión D\_REPETIDOR que indica si un alumno es extranjero o no.

- **D\_SEXO:** Dimensión que indica el sexo de un alumno. Posee 2 campos. El primero almacena una “H” si es hombre o una “M” si es mujer. El segundo es un descriptivo del primero.
- **D\_TIPO\_CENTRO:** Dimensión que indica si un centro es privado o público.
- **D\_MISMO\_MUNICIPIO:** Dimensión que indica si el municipio de residencia de un alumno es igual al municipio del centro.

La tabla de hechos contendrá una métrica que será la nota media del alumno.

### 5.1. Tablas de dimensiones

Para la creación de las transformaciones, se ha decidido incluir en cada una de ellas, un paso de entrada de tabla, que establecería una conexión con la base de datos de RACIMA y con la tabla objetivo. El paso de entrada de tabla irá unido a otro que será una salida de tabla, la cuál será una tabla de hechos o de dimensiones. Previamente a la ejecución de la transformación, se ha creado un índice en la tabla de dimensiones correspondiente a la transformación con el objetivo de unir mediante dicho índice la tabla de dimensiones con la de hechos. Una vez creado el índice, se ha procedido a ejecutar la transformación para cargar los datos de la tabla origen en la tabla de dimensiones. A continuación se muestran los pasos de la transformación (Figura 5.1) y de la configuración del paso (Figura 5.2 y Figura 5.3) de la dimensión **D\_MUNICIPIOS**. Los pasos seguidos para esta dimensión son idénticos en el resto de dimensiones.



*Figura 5.1. Pasos de la transformación de la dimensión D\_SEXO*

Entrada Tabla

Nombre paso: Entrada Tabla Sexo

Conexión: racima

SQL

```
select distinct L_SEXO, case when L_SEXO='H' then 'Hombre'
else 'Mujer' end as DESC_SEXO
from tlalumnos
```

Line 3 Column 14

Enable lazy conversion ☐

Reemplazar variables en script? ☐

Insertar datos del paso

Ejecutar para cada fila? ☐

Limitar tamaño 0

Help Vale Previsualizar Cancelar

Figura 5.2. Configuración del paso de Entrada de tabla para la tabla D\_SEXO

Salida de Tabla

Nombre paso: Salida Tabla Sexo

Conexión: racima

Esquema destino: RACIMA\_GESCEN

Tabla destino: D\_SEXO

Tamaño transacción (commit): 1000

Vaciar tabla ☒

Ignorar errores de inserción ☐

Specify database fields ☐

Main options Database fields

Repartir información en varias tablas ☐

Campo de partición

Particionar información por mes ☒

Particionar información por días ☐

Utilizar actualización por lotes para inserciones ☒

El nombre de la tabla está definido en un campo? ☐

Campo que contiene el nombre de la tabla:

Almacena el campo con el nombre de tabla ☒

Incluye clave auto-generada ☐

Nombre del campo clave auto-generada

Help Vale Cancelar SQL

Figura 5.3. Configuración del paso de Salida de tabla para la tabla D\_SEXO

Cabe destacar que en el paso de Salida de tabla, está marcada la opción de “Vaciar tabla”. Esto se ha realizado así para que cada vez que se ejecute la transformación, Pentaho primero va a vaciar la tabla de dimensión y, posteriormente, cargará los nuevos datos de la tabla origen en ella.

## 5.2. Tablas de hechos

Para la creación de la tabla de hechos, se ha optado por crear una tabla llamada H\_ALUMNOS que representa todos los aspectos importantes de los alumnos obtenidos durante el estudio de los alumnos de 4º de la E.S.O.

La tabla contiene el ID del alumno, el curso, un flag que indica si es repetidor o no, un flag para indicar si ese alumno tiene conductas contrarias o no, un identificador de grupo de faltas de asistencia al que pertenece según las faltas de asistencia registradas para dicho alumno, la nota media del curso, un flag para indicar si es de nacionalidad española o extranjera, un flag indicando su sexo, el tipo de centro al que asiste y si su municipio de residencia es el mismo que el municipio del centro.

La realización de esta tabla ha tenido mayor complejidad que la de las dimensiones. A continuación se muestra la consulta utilizada para esta tarea en el paso **entrada de tabla** de Pentaho.

```
select cur.C_ANNO as curso, mat.l_repite as repite,  
       case when exists (select distinct alu.x_alumno  
                           from tlmatalu alu join  
                           tluconcon alucon on  
                           alu.x_matricula=alucon.x_matricula) then 'S' else 'N' end  
       as conductas_contrarias,  
       case when (select count (*) as faltas from tlmatalu alu join tlfalasialu falasi on  
                   alu.x_matricula=falasi.x_matricula)=0 then 0  
              when (select count (*) as faltas from tlmatalu alu join tlfalasialu  
                   falasi on alu.x_matricula=falasi.x_matricula) > 0 and (select count (*)  
                   as faltas from tlmatalu alu join tlfalasialu falasi on  
                   alu.x_matricula=falasi.x_matricula)<=99 then 1
```



```

        when (select count (*) as faltas from tlmatalu alu join tlfalasialu
        falasi on alu.x_matricula=falasi.x_matricula)>99 and (select count (*)
        as faltas from tlmatalu alu join tlfalasialu falasi on
        alu.x_matricula=falasi.x_matricula)<=199 then 2
        when (select count (*) as faltas from tlmatalu alu join tlfalasialu
        falasi on alu.x_matricula=falasi.x_matricula)>199 and (select count (*)
        as faltas from tlmatalu alu join tlfalasialu falasi on
        alu.x_matricula=falasi.x_matricula)<=299 then 3
        when (select count (*) as faltas from tlmatalu alu join tlfalasialu
        falasi on alu.x_matricula=falasi.x_matricula)>299 and (select count (*)
        as faltas from tlmatalu alu join tlfalasialu falasi on
        alu.x_matricula=falasi.x_matricula)<=399 then 4
        else 5
    end as faltas_asistencia,
    nota.nota_media,
    case when alu.c_pais=034 then 1 else 2 end as nacionalidad,
    case when alu.l_sexo='H' then 'H' when alu.l_sexo='M' then 'M' else 'NC' end as
    sexo,
    cen.x_tipo as tipo_centro,
    case when (alu.c_municipio!=dat.c_municipio) then 'distinto municipio' else
    'mismo municipio' end as mismo_municipio,
    alu.x_alumno as alumno

from (select alu.x_alumno, avg(cal.n_numero) as nota_media from tlmatalu alu join
tmatmaterialu mat on mat.x_matricula=alu.x_matricula
join tcalificaciones cal on mat.x_califica=cal.x_califica group by alu.x_alumno) nota
join tlalumnos alu on nota.x_alumno=alu.x_alumno
join tlmatalu mat on alu.x_alumno=mat.x_alumno
join tlcursosaca cur on mat.c_anno=cur.c_anno
join tmatmaterialu matr on mat.x_matricula=matr.x_matricula
join tcalificaciones cal on matr.x_califica=cal.x_califica
join tldatoscen dat on mat.x_centro=dat.x_centro
join tlcensos cen on dat.x_centro=cen.x_centro;

```

A continuación, en la Figura 5.4 y en la Figura 5.5, se muestra el resultado de la ejecución de la transformación anterior:

[illegible]

**Figura 5.4.** Primer resultado de la ejecución de la transformación para la tabla H\_ALUMNOS

NACIONALIDAD	SEXO	TIPO_CENTRO	MISMO_MUNICIPIO	ALUMNO
2 M		1 mismo municipio		3216559
2 M		1 mismo municipio		3216559
1 M		1 distinto municipio		3216561
1 M		1 distinto municipio		3216561
1 M		1 distinto municipio		3216561
1 M		1 distinto municipio		3216561
1 M		1 distinto municipio		3216249
1 M		1 distinto municipio		3216249

**Figura 5.5.** Segundo resultado de la ejecución de la transformación para la tabla H\_ALUMNOS

Tras este último paso, se ha conseguido construir un Data Warehouse para la realización de un cubo. Un cubo es una estructura de datos organizada mediante jerarquías. Cada indicador se puede evaluar en cualquiera de los niveles de las jerarquías.

## 6. Creación de un cubo

Con el objetivo de poder visualizar los datos y construir un pequeño cuadro de mando, se ha decidido construir un cubo. El código del cubo irá en un archivo XML. Para conseguir la visualización, habrá que hacer uso del servidor de Pentaho, que se obtiene al descargar Pentaho Spoon.

A continuación se explica el proceso de creación del archivo XML utilizado para la construcción del cubo.

- **Dimensiones**

A continuación se van a explicar los elementos más importantes de una dimensión debido a que las demás dimensiones se construyen de la misma forma.

```
<Schema name="RACIMA" measuresCaption="Medidas">
  <Dimension highCardinality="false" name="CURSO_ACADEMICO">
    <Hierarchy hasAll="true" allMemberName="Total cursos" primaryKey="X_CURSO">
      <Table name="D_CURSO">
        </Table>
        <Level name="cursos" column="X_CURSO" nameColumn="DESC_CURSO"
ordinalColumn="X_CURSO" type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
          </Level>
        </Hierarchy>
      </Dimension>
    ...
  </Schema>
```

Lo primero que se muestra son las etiquetas <Schema> para indicar el inicio y fin del esquema del documento. Seguidamente se abre la etiqueta <Dimension> para indicar que el contenido de esa etiqueta es la definición de la dimensión (D\_CURSO en este caso), en el que se le da un nombre para poder referenciar a esta dimensión más adelante (CURSO\_ACADEMICO). En la siguiente línea se indica el campo de la tabla que es la clave primaria (X\_CURSO). A

continuación se indica la tabla origen de la que obtiene los datos (la dimensión D\_CURSO).

La línea siguiente indica la clave primaria y le otorga un alias (DESC\_CURSO), con el que referenciar a dicha columna con posterioridad. De tratarse de una dimensión jerárquica, habría en este punto más etiquetas <Level>.

### ● Hechos

```
<Cube name="ALUMNOS" cache="true" enabled="true">
  <Table name="H_ALUMNOS">
    </Table>
    <DimensionUsage source="CURSO_ACADEMICO" name="CURSO_ACADEMICO"
foreignKey="X_CURSO" highCardinality="false">
      </DimensionUsage>
  ...
  <Measure name="Alumnos" column="ALUMNO" formatString="#,###" aggregator="distinct-
count">
    </Measure>
</Cube>
</Schema>
```

Cada tabla de hechos va dentro de las etiquetas <Cube> junto con el nombre que se le ha proporcionado. Seguidamente se le indica la tabla de la que tiene que obtener los datos (H\_ALUMNOS). Las etiquetas <DimensionUsage> se utilizan para indicar las relaciones con las dimensiones. En este caso, se indica que el origen es la dimensión llamada CURSO\_ACADEMICO y que la clave foránea es X\_CURSO.

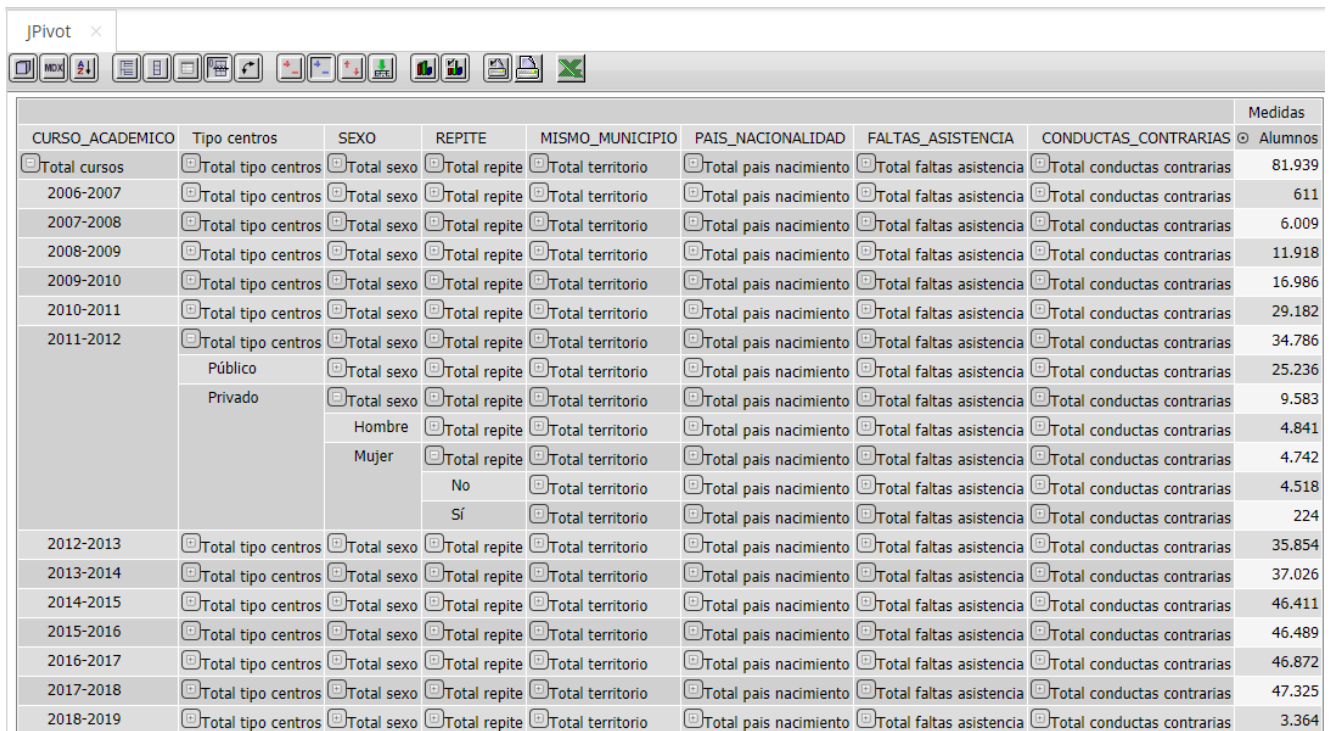
Las etiquetas <Measure> encierran las medidas de la tabla de hechos. En este caso, se escoge la columna Alumnos, se le indica el formato numérico que se va a utilizar y la función de agregación (distinct-count).

Tras este último paso, ya se tiene el código del cubo terminado. Para poder visualizar los datos, es necesario utilizar Pentaho-Server. Una vez se ha logueado como administrador, hay que crear la conexión a la base de datos de RACIMA en la opción **Manage data sources** y crear la conexión con la opción

**New connection.** Seguidamente, se selecciona la opción **import analysis** para importar el fichero XML que contiene el cubo.

Pentaho-Server tiene un plugin llamado JPivot que sirve para visualizar los datos de un cubo. JPivot es la herramienta que se utilizará para visualizar los datos.

En la Figura 6.1, se muestra un cuadro de mando simple utilizando el cubo creado en el paso anterior.



CURSO_ACADEMICO	Tipo centros	SEXO	REPITE	MISMO_MUNICIPIO	PAIS_NACIONALIDAD	FALTAS_ASISTENCIA	CONDUCTAS_CONTRARIAS	Medidas
Total cursos	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	Alumnos 81.939
2006-2007	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	611
2007-2008	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	6.009
2008-2009	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	11.918
2009-2010	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	16.986
2010-2011	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	29.182
2011-2012	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	34.786
	Público	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	25.236
	Privado	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	9.583
		Hombre	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	4.841
		Mujer	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	4.742
		No	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	4.518
		Sí	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	224
2012-2013	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	35.854
2013-2014	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	37.026
2014-2015	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	46.411
2015-2016	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	46.489
2016-2017	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	46.872
2017-2018	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	47.325
2018-2019	Total tipo centros	Total sexo	Total repite	Total territorio	Total pais nacimiento	Total faltas asistencia	Total conductas contrarias	3.364

**Figura 6.1.** Cuadro de mando en Pentaho-Server

## 7. Comparativa entre el tiempo estimado y el real

Descripción tareas	
Nombre paquete	Descripción
1.1. Planificación	Definición de las tareas a realizar, la forma de proceder y las horas estimadas.
1.2. Alcance y requisitos	Especificación de cuál es el objetivo y lo que se pretende conseguir y lo que no.
1.3. Seguimiento y control	Recoge los cambios que ha ido sufriendo el proyecto a lo largo de su realización e incluye comparaciones con lo planificado.
1.4. Reuniones	Información sobre las reuniones mantenidas tanto con el cliente como con los tutores.
<b>Tiempo estimado</b>	<b>70 horas</b>
<b>Tiempo real</b>	<b>53 horas</b>
2.1. Pentaho	Lectura de manuales sobre Pentaho además de su instalación y la realización de pruebas siguiendo tutoriales.
2.2. BD Racima	Instalación de la base de datos y estudio de las tablas que la componen y de su estructura, tomando anotaciones para aligerar futuras consultas sobre ella.
<b>Tiempo estimado</b>	<b>90 horas</b>
<b>Tiempo real</b>	<b>90 horas</b>
3.1. Diseño	Diseños realizados después del estudio de la base de datos para encontrar posibles hechos de interés para el proyecto.
3.2. Implementación	Construcción de los hechos en formato de estrella utilizando la herramienta Pentaho siguiendo los diseños que finalmente se hayan seleccionado.

3.3. Documentación	Información sobre las tablas de la base de datos Racima. Incluye el nombre de la tabla, lo que representa y los campos considerados más importantes.
<b>Tiempo estimado</b>	<b>90 horas</b>
<b>Tiempo real</b>	<b>110 horas</b>
4.1. Memoria	Redacción de la memoria del proyecto.
4.2. Defensa	Presentación del proyecto acompañado de un powerpoint.
<b>Tiempo estimado</b>	<b>50 horas</b>
<b>Tiempo real</b>	<b>52 horas</b>
<b>Tiempo total estimado</b>	<b>300 horas</b>
<b>Tiempo total real</b>	<b>305 horas</b>

Las principales causas de los descuadres en la tabla anterior se deben a las dificultades encontradas a la hora de reunirse con el cliente, a problemas con la importación de la base de datos y a problemas encontrados durante la fase de implementación.

## 8. Conclusiones

En este Trabajo Fin de Grado, se ha realizado un estudio de las causas del fracaso escolar en la Comunidad de La Rioja, mediante un ejercicio de ETL sobre los datos aportados por el cliente que se encuentran almacenados en la base de datos RACIMA.

Para llevar a cabo este proyecto y conseguir el objetivo, primeramente se ha estudiado completamente la base de datos RACIMA, entendiendo el significado de todas sus tablas.

Posteriormente, se ha diseñado un Data Warehouse con una tabla de hechos, a partir de las dimensiones detectadas durante la fase de estudio,

Por último, también se ha realizado un ejercicio de ETL sobre la base de datos RACIMA para pasar los datos de la base de datos original al nuevo diseño, creando así las tablas de dimensiones y hechos.

Como conclusiones particulares, destacar que tras analizar la base de datos RACIMA, se han detectado los siguientes factores significativos relacionados con el fracaso escolar:

	Alumnos con fracaso escolar	Alumnos sin fracaso escolar
Repite curso	86,67 %	29,46 %
Conductas contrarias	51,51 %	37,51 %
Faltas de asistencia	89,70 %	82,92 %
Media notas	5,11	6,85
Alumnos españoles	6,10 %	93,18 %
Alumnos extranjeros	24,04 %	71,24 %
Hombres	10,04 %	88,52 %
Mujeres	7,05 %	91,88 %
Necesitan moverse de municipio	34,08 %	3,64 %

Finalmente, agradecer a Francisco Martínez Martín y Óscar Alonso Hernández por colaborar y ayudar en todo momento y sobre todo, estar a mi entera disposición.



## 9. Bibliografía

- **Descarga de Pentaho Data integration:**  
<https://sourceforge.net/projects/pentaho/files/Data%20Integration/>
- **Descarga de SQL Developer**  
<https://www.oracle.com/technetwork/developer-tools/sql-developer/downloads/index.html>
- **Tutorial Pentaho:**  
<http://holowczak.com/building-etl-transformations-in-pentaho-data-integration-kettle/5/>
- **Ejemplo ejercicio práctico BI:**  
<https://www.cs.buap.mx/~zacarias/FZF/SAKILA-Ejercicio.pdf>
- **Información sobre RACIMA:**  
<https://www.larioja.org/tecnologia-comunica/es/proyectos-dgtic/servicios-informaticos-areas-gobierno/educacion/racima>
- **Definición de cubo de datos:**  
<https://www.businessintelligence.info/definiciones/cubos-olap.html>